# Weak Identification Robust Tests in an Instrumental Quantile Model

Sung Jae Jun[1]

Brown University

Department of Economics

Version as of September 2005

(Job Market Paper)

**Abstract**

We consider the linear instrumental quantile model proposed by Chernozhukov and Hansen (2001, 2005a, 2005b). Since it is never clear for what quantile effects the given instruments are most informative, we develop a testing procedure that is robust to identification quality in the GMM framework. In order to reduce the computational burden, a multi-step approach is taken, and a two-step Anderson-Rubin (AR) statistic is considered. We then propose a three-step orthogonal decomposition of the AR statistic, where the null distribution of each component does not depend on the assumption of a full rank of the Jacobian. Power experiments are conducted, and inferences on returns to schooling using the Angrist and Krueger data are considered as an empirical example. Although returns to schooling for the upper quantiles seem to be quite constant, the robust confidence sets for the lower quantiles are so wide that they still remain imprecise, which differs from the results in Lee (2004) and Chernozhukov and Hansen (2005b).

*Key Words*: Quantile Regression, Instruments, GMM, Weak Identification

[1] Department of Economics, Box B, Brown University, Providence, RI 02912, USA, e-mail: sjun@brown.edu

I am grateful to F. Kleibergen, T. Lancaster, S. Mavroeidis, and my colleagues in the department for many helpful comments and discussions. All remaining errors are my own.

# 1 Introduction

Quantile regression models are useful to analyze the impact of treatment variables on the distribution of outcomes and have been used in many economic applications (e.g., Buchinsky (1994), Abadie (1995), among others). However, in many cases, the treatment variables are endogenous, and conventional quantile regression is inappropriate for causal or structural analysis. To overcome this difficulty, several modified models have been proposed (see e.g., Amemiya (1975), Abadie, Angrist and Imbens (2002), Chernozhukov and Hansen (2001, 2004, 2005a, 2005b), Imbens and Newey (2002), Chesher (2003), Ma and Koenker (2004), Lee (2004)). Chernozhukov and Hansen take an instrumental variable approach. In constrast to Abadie et al. (2002), Chernozhukov and Hansen's model allows for a non-binary endogenous variable. It is also more general than Amemiya (1975), which is a location model that does not allow for treatment effect heterogeneity. More importantly, this model can be cast in the GMM framework in spite of its computational difficulty.

We take Chernozhukov and Hansen's approach to endogenous quantile models and consider testing the parameter of an endogenous variable without making the identification assumption in the GMM framework. Removing the identification assumption is particularly interesting in quantile analysis because defending instruments is not as easy as in a mean regression model, and the identification condition depends on what quantile effects are considered. It is never clear for what quantile effects the given instruments are most informative. In particular, when instruments are "weak," the identification of some quantile effects can be relatively harder than that of others. Since the mean effect is the integral of all the quantile effects, the identification of the mean effect is more demanding than that of a particular quantile effect.

As is discussed in Dufour (1997) and Stock and Wright (2000), among others, it is well known that the normal approximation of the GMM estimator can be extremely poor without the strong identification assumption, which is not always obvious for verification in practice. One might want to pretest the quality of the identification (Hahn and Hausman (2002), Wright (2002)), but this is not thought of as a desirable approach because the consequential inference should be conditional on the pretest. For this reason, many efforts have been made to develop testing procedures that are robust to the potential identification difficulty (Stock and Wright (2000), Moreira (2003), Kleibergen (2002, 2004, 2005)). So far, these efforts seem to be quite successful. In particular, Kleibergen (2005) shows that the score statistic based on the continuous-updating GMM is robust to weak values of *the expectation of the Jacobian* by assuming the differentiability of the GMM objective function.[2] The differentiability assumption is restrictive because it rules out interesting models such as the instrumental quantile model considered in this paper.

The non-smoothness of the objective function causes several problems: the computation becomes harder and the score statistic is not available. To avoid a high dimensional grid search, the coefficients of the exogenous variables are concentrated out under the null hypothesis of the structural parameter. This approach significantly reduces the computational burden and results in a straightforward two-step Anderson-Rubin statistic ($AR$ statistic), which is a joint test of the location of the parameter and the overidentifying restrictions (Anderson and Rubin (1949), Stock and Wright (2000), Honore and Hu (2004)).[3] We then show how to obtain a three-step orthogonal decomposition of the $AR$ statistic. Simulation experiments demonstrate the power improvement achieved by this decomposition.

We then apply this method to the study of returns to schooling using the Angrist and Krueger data (e.g., Angrist and Krueger (1991) and Angrist, Imbens, and Krueger (1999)). Confidence sets of returns to schooling for various quantiles are obtained by inverting the two-step AR test and its orthogonal decomposition. Although returns to schooling for the upper quantiles seem to be quite constant, the robust confidence sets for the lower quantiles are so wide that they still remain obscure, which differs from the results in Lee (2004) and Chernozhukov and Hansen (2005b).

The paper is organized as follows. Section 2 presents a brief overview of the instrumental quantile model. The weak identification and computational issues are discussed in the next section. Section 4 considers testing the parameter of the endogenous variable, and section 5 extends it. The simulation results are then shown, and the example of returns to schooling is discussed in section 7. Conclusions are presented in the last section.

---

[2] Local identification in general GMM requires full rank of *the Jacobian of the expectation*, which is well defined even if *the expecation of the Jacobian* is not (see e.g., Pakes and Pollard (1989)).

[3] In the recent revised version of Chernozhukov and Hansen (2001), they propose a dual inference that is robust to the instrument quality (Chernozhukov and Hansen (2004)). This idea is in line with the $AR$ statistic, which is a joint test of the location and the overidentifying restrictions. In fact, it is easy to show that this idea leads to the $AR$ statistic in the standard linear mean regression model.

## 2 Basic Setup

This section reviews the instrumental quantile model by Chernozhukov and Hansen (2001, 2004, 2005a, 2005b). A single endogenous variable $D$ whose support is $\mathcal{D} \subset \mathcal{R}$ is considered. Counterfactual outcomes are indexed against the treatment and denoted $Y_d$ where $d \in \mathcal{D}$, and the exogenous variables whose support is $\mathcal{X} \subset \mathcal{R}^k$, are denoted $X$. The object of interest is the quantile function of $Y_d$ conditional on $X = x$ and it is denoted by $q(d, x, \tau)$. It is assumed that $q(d, x, \tau)$ is given by (or well approximated by) a linear form for each $\tau \in (0, 1)$. That is, given $X = x$,

$$q(d, x, \tau) = d\alpha(\tau) + x'\beta(\tau) \text{ where } 0 < \tau < 1 \tag{2.1}$$

Given the actual treatment $D$, the observed outcome is $\overline{Y} \equiv Y_D$. In other words, only the $D^{th}$ component of $\{Y_d\}_{d \in \mathcal{D}}$ is observed. The parameter of interest is $\alpha(\tau)$ for a given $\tau$. It is the causal effect of the exogenous change of the treatment on the $\tau$-th quantile of the counterfactual outcomes. For the sake of simplicity, it is assumed that all $Y_d$'s are continuous random variables. Hence, $\overline{Y} = Y_D$ is a continuous random variable and $q(d, x, \tau) = d\alpha(\tau) + x'\beta(\tau)$ is strictly increasing in $\tau$ for every $d \in \mathcal{D}$ and $x \in \mathcal{X}$. For representation, it is assumed that $D = \delta(Z, X, V)$, where $V$ is a random disturbance term and that $Y_d = d\alpha(U) + x'\beta(U)$ given $X = x$, where $U$ is a uniformly distributed random variable.[4] It is also assumed that $Z$ and $X$ are independent of $U$. The correlation between $U$ and $V$ (or $U$ and $D$) causes the endogeneity problem.

The statistical implication of this setup is that $\Pr(\overline{Y} \leq D\alpha(\tau) + X'\beta(\tau) \mid Z, X) = \tau$.[5] This is the conditional moment restriction considered in this paper. Note that $\Pr(\overline{Y} \leq D\alpha(\tau) + X'\beta(\tau) \mid D, X) \neq \tau$ due to the correlation between $U$ and $D$, which makes traditional quantile regression inappropriate (see e.g., Buchinsky (1998)).

This model can be rewritten as follows, which may be more conventional and familiar.

$$\begin{aligned} \overline{Y} &= D\alpha(\tau) + X'\beta(\tau) + \eta \\ D &= \delta(Z, X, V) \end{aligned} \tag{2.2}$$

[4] For the representation of $Y_d$, a uniform random variable $U_d$ is needed for each $d \in \mathcal{D}$. It is assumed that all $U_d$'s collapse to a common random variable $U$. This is a stronger assumption than what we actually need (see Chernozhukov and Hansen (2001)).

[5] The intuition for this is simple. $\Pr(q(D, X, U) \leq q(D, X, \tau) \mid Z, X) = \Pr(U \leq \tau \mid Z, X) = \Pr(U \leq \tau) = \tau$ (see Chernozhukov and Hansen (2001)).

where $\eta = \eta(D, X, U, \tau)$ and $\Pr(\eta \leq 0 \mid Z, X) = \tau$. This is more general than the conventional location model in the sense that the treatment effect heterogeneity is allowed for different $\tau$s. The main goal of this paper is to develop a procedure to test $H_0 : \alpha(\tau) = \alpha_0$ without making the identification assumption in the GMM framework. Hence, $\alpha(\tau)$ is assumed to be known under the null hypothesis throughout the paper. A 95% confidence set can be obtained by inverting a hypothesis test of size 5%. The point estimation of $\alpha(\tau)$ is not pursued because it is meaningless without the identification assumption. Giving up the point estimation, a confidence set can still be obtained even though it may be the whole parameter space when the data are totally uninformative. For further discussion, see Anderson and Rubin (1949), Dufour (1997), Stock and Wright (2000), Kleibergen (2002, 2004, 2005), and Moreira (2004), among others.

# 3 GMM and the Weak Identification

The previous discussion provides a conditional moment restriction: $E(1\{\overline{Y} \leq D\alpha + X'\beta\} \mid Z, X) = \Pr(\overline{Y} \leq D\alpha + X'\beta \mid Z, X) = \tau$ when $\alpha = \alpha(\tau), \beta = \beta(\tau)$, where $1\{\cdot\}$ is the standard indicator function.[6] Therefore, an unconditional moment restriction can be obtained as follows.

$$\begin{aligned} m(\alpha, \beta) &= \begin{bmatrix} m_1(\alpha, \beta) \\ m_2(\alpha, \beta) \end{bmatrix} \\ &= E(Q_X(1\{\overline{Y} \leq D\alpha + X'\beta\} - \tau)) = 0 \text{ when } \alpha = \alpha(\tau), \beta = \beta(\tau). \end{aligned} \tag{3.1}$$

where $Q_X = \begin{bmatrix} X & W' \end{bmatrix}'$ and $W = W(X, Z)$. The parameter of interest is the structural parameter $\alpha(\tau)$.

Although it is conceptually straightforward to obtain the GMM estimator from these moment conditions, there are several difficulties. First, the indicator function is not a smooth function and the optimization requires a grid search over a potentially high dimensional space. Second, the consistency and asymptotic normality of the GMM estimator requires that $m(\alpha, \beta) = 0$ *only* when $\alpha = \alpha(\tau), \beta = \beta(\tau)$ and that the Jacobian, $\frac{\partial m}{\partial(\alpha, \beta')} \mid_{\alpha=\alpha(\tau), \beta=\beta(\tau)}$, has a full column rank. These are the global and the local identification conditions, respectively. When the Jacobian is equal to zero, the parameters are not (locally) identified, and the asymptotic properties of the GMM estimator break down. This lack of identification implies that identifying the parameters can be extremely

[6] For the sake of simplicity, the qualifyer of "almost surely" will be suppressed throughout the paper when it is clear from the context.

difficult when the Jacobian is not zero but arbitrarily close to zero (see Dufour (1997)). Stock and Wright (2000) propose an alternative asymptotic tool to capture this phenomenon by setting $\frac{\partial m}{\partial(\alpha,\beta')}\mid_{\alpha=\alpha(\tau),\beta=\beta(\tau)}$ to be local to zero. They show that the GMM estimator is not consistent and not asymptotically normal under this alternative asymptotics, which they call the weak identification asymptotics. The condition of the full column rank of the Jacobian is, however, just an assumption and it is not clear how to check it in practice.[7]

The fact that weak values of $\frac{\partial m}{\partial(\alpha,\beta')}\mid_{\alpha=\alpha(\tau),\beta=\beta(\tau)}$ cause a problem has an important practical implication. Since $(\alpha(\tau_1),\beta(\tau_1)')\neq(\alpha(\tau_2),\beta(\tau_2)')$ for $\tau_1\neq\tau_2$, the Jacobian depends on what quantile effect is examined. Some quantiles may be better identified and the given instruments can be more informative about one quantile than another. Since the mean effect is the integral of all the quantile effects, the identification of the mean effect is more demanding than that of a quantile effect. Therefore, it is particularly interesting to develop a robust inference method in the quantile regression context.

Kleibergen (2005) shows how to conduct robust inference in GMM against the potential identification difficulty by assuming the differentiability of the function involved in the moment conditions. He shows that a score statistic from the continuous updating objective function is robust to weak values of the expectation of the Jacobian. This approach does not apply to the current model because of the non-smoothness of the indicator function. It should also be noted that what matters for identification is not *the expectation of the Jacobian* but *the Jacobian of the expectation*. Although Kleibergen's method is not directly applicable, it gives an important insight. The key point is not having the score of the GMM objective function but estimating the Jacobian in such a way that it is asymptotically independent of the $\sqrt{n}$ times sample analog of $m(\alpha(\tau),\beta(\tau))$ as will be shown later.

We give some remarks on the smoothed GMM approach (see e.g., Horowitz (1992) and MaCurdy and Timmins (2001)). One might think that the standard robust statistics can be applied by directly smoothing the indicator with a suitable distribution function (cdf type kernel). However, this approach is not pursued in this paper because it makes the problem unnecessarily complicated

[7] In the linear mean location model, it is relatively easier to defend this condition because it is equivalent to the sufficiently strong correlation between the endogenous variable and the instruments. However, it is not clear how to check this condition in general nonlinear models. Honore and Hu (2004) write, "it is not clear exactly which (easily interpretable) assumptions would lead to this being true in general."

due to the fact that the bias and under-smoothing issue should be dealt with from the first step of defining the $AR$ statistic.

### 3.1 Intuitive Discussion

Although the main interest of the paper is in $\alpha(\tau)$, the joint test of $\alpha(\tau)$ and $\beta(\tau)$ will be considered first for an intuitive discussion. Some computational issues in concentrating out $\beta(\tau)$ will be pointed out and a practical strategy for testing $\alpha(\tau)$ will be discussed. In the following, $\alpha(\tau)$ and $\beta(\tau)$ are regarded as the hypothesized values under the null.

**Assumption 1** *$\eta \equiv \overline{Y} - D\alpha(\tau) - X'\beta(\tau)$ has a well defined density conditional on $D, Z, X$, $f_\eta(s \mid D, Z, X)$ and it is bounded away from 0 at $s = 0$.*

Consider $m(\alpha, \beta) + \tau E(Q_X) = E(Q_X 1\{\overline{Y} \leq D\alpha + X'\beta\}) = E(Q_X 1\{\eta \leq D(\alpha - \alpha(\tau)) + X'(\beta - \beta(\tau))\})$ from the definition of $\eta$. By the law of iterated expectation, it is easy to see that

$$\begin{aligned} m(\alpha, \beta) + \tau E(Q_X) &= E(Q_X E(1\{\eta \leq D(\alpha - \alpha(\tau)) + X'(\beta - \beta(\tau))\} \mid D, Z, X)) \quad (3.1.1) \\ &= E(Q_X F_\eta(D(\alpha - \alpha(\tau)) + X'(\beta - \beta(\tau)) \mid D, Z, X)) \end{aligned}$$

where $F_\eta(\cdot \mid \cdot)$ is the conditional distribution function of $\eta$. The second equality is due to the fact that the conditional expectation of an indicator is a conditional probability. From (3.1.1), it can be seen that

$$\begin{aligned} J(\alpha(\tau), \beta(\tau)) &\equiv \frac{\partial m(\alpha, \beta)}{\partial(\alpha, \beta')} |_{\alpha=\alpha(\tau), \beta=\beta(\tau)} \\ &= \begin{bmatrix} E(XDf_\eta(0 \mid D, Z, X)) & E(XX'f_\eta(0 \mid D, Z, X)) \\ E(WDf_\eta(0 \mid D, Z, X)) & E(WX'f_\eta(0 \mid D, Z, X)) \end{bmatrix} \end{aligned}$$

Notice that if the conditional density $f_\eta(s \mid D, Z, X)$ were known, we would obtain a straightforward estimator for $J(\alpha(\tau), \beta(\tau))$, which could be used to obtain a robust test statistic. Specifically, let

$$M_i \equiv \begin{bmatrix} X_i D_i f_\eta(0 \mid D_i, Z_i, X_i) & X_i X_i' f_\eta(0 \mid D_i, Z_i, X_i) \\ W_i D_i f_\eta(0 \mid D_i, Z_i, X_i) & W_i X_i' f_\eta(0 \mid D_i, Z_i, X_i) \end{bmatrix}.$$

Then, $\frac{1}{n}\sum_{i=1}^{n} M_i \xrightarrow{p} J(\alpha(\tau), \beta(\tau))$ under mild regularity conditions. Following Kleibergen (2002, 2005), we make the following assumption.

**Assumption 2** *The data are i.i.d. and a central limit theorem (CLT) holds under the null. In particular, we have*

$$\begin{bmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Q_{Xi}(1\{\overline{Y}_i \leq D_i\alpha(\tau) + X_i'\beta(\tau)\} - \tau) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(vec(M_i) - vec(J(\alpha(\tau),\beta(\tau))) \end{bmatrix} \xrightarrow{d} N(0, \begin{bmatrix} \widetilde{V}_{11} & \widetilde{V}_{12} \\ \widetilde{V}_{21} & \widetilde{V}_{22} \end{bmatrix})$$

*where $\widetilde{V}_{11}$ is positive definite.*[8]

**Lemma 1** *Let $\widehat{m} \equiv \frac{1}{n}\sum_{i=1}^{n} Q_{Xi}(1\{\overline{Y}_i \leq D_i\alpha(\tau) + X'\beta(\tau)\} - \tau)$ and $vec(\widehat{J}) \equiv \frac{1}{n}\sum_{i=1}^{n} vec(M_i) - \widetilde{V}_{21}\widetilde{V}_{11}^{-1}\widehat{m}$.*

*Then,* $$\begin{bmatrix} \sqrt{n}\widehat{m} \\ \sqrt{n}(vec(\widehat{J}) - vec(J(\alpha(\tau),\beta(\tau))) \end{bmatrix} \Rightarrow \begin{bmatrix} \widetilde{\Psi}_1 \\ vec(\widetilde{\Psi}_2) \end{bmatrix} \sim N(0, \begin{bmatrix} \widetilde{V}_{11} & 0 \\ 0 & \widetilde{V}_{22} - \widetilde{V}_{21}\widetilde{V}_{11}^{-1}\widetilde{V}_{12} \end{bmatrix}).$$

**Proof.** Results directly from Assumption 2. ■

The most straighforward way of testing is to use the distribution of $\sqrt{n}\widehat{m}$, which results in the famous $AR$ statistic:

$$AR^*(\alpha(\tau),\beta(\tau)) \equiv (\sqrt{n}\widehat{m})'\widetilde{V}_{11}^{-1}(\sqrt{n}\widehat{m}) \xrightarrow{d} \chi^2(\dim(Q_X))$$

regardless of identification. Although $AR^*(\alpha(\tau),\beta(\tau))$ is an easy method of testing, it wastes degrees of freedom when the model is highly overidentified. Therefore, as Kleibergen (2002, 2005) suggests, its power property can be improved by projecting $\sqrt{n}\widehat{m}$ on the space determined by the Jacobian part. Note that $\widehat{J}$ is estimating $J(\alpha(\tau),\beta(\tau))$ in the sense that $\widehat{J} \xrightarrow{p} J(\alpha(\tau),\beta(\tau))$ and that it is asymptotically independent of $\sqrt{n}\widehat{m}$. Its independence is crucial for the following theorem.

**Theorem 1** *Let $S_n^* \equiv (\sqrt{n}\widehat{J})'\widetilde{V}_{11}^{-1}\sqrt{n}\widehat{m}$. Then, $K_n^*(\alpha(\tau),\beta(\tau)) \equiv S_n^{*\prime}((\sqrt{n}\widehat{J})'\widetilde{V}_{11}^{-1}(\sqrt{n}\widehat{J}))^{-1}S_n^* \xrightarrow{d} \chi^2(1+k)$ when $J(\alpha(\tau),\beta(\tau))$ is (i) 0, (ii) $O(\frac{1}{\sqrt{n}})$ or (iii) of full rank, where $k$ is the dimension of $\beta(\tau)$.*

**Proof.** See appendix. ■

Since $f_\eta(0 \mid D, Z, X)$ is not known, $K_n^*(\alpha(\tau),\beta(\tau))$ is infeasible. This discussion, however, shows how the robust statistics can be obtained when the GMM objective function is not differentiable.

[8] The dependency of the variance on $\alpha(\tau),\beta(\tau)$ is suppressed for the sake of simplicity. Throughout this paper, all the variances will be continuous updating ones that depend on the null hypothesis.

The score of the objective function is not available but the score statistic can be mimicked by using the score of the limit function. Note also that $K_n^*(\alpha(\tau),\beta(\tau)) = (\sqrt{n}\widehat{m})'V^{-\frac{1}{2}}P_{\widetilde{V}_{11}^{-\frac{1}{2}}\widehat{J}}V^{-\frac{1}{2}}(\sqrt{n}\widehat{m})$, where $P_{\widetilde{V}_{11}^{-\frac{1}{2}}\widehat{J}}$ is a projection onto the column space of $\widetilde{V}_{11}^{-\frac{1}{2}}\widehat{J}$. Therefore, $K_n^*(\alpha(\tau),\beta(\tau))$ also gives the orthogonal decomposition of the Anderson-Rubin statistic.[9]

### 3.2 Computational Issue and Concentrated Moment Condition: Two-Step AR Statistic

Recall that the main interest of this paper is in testing the structural parameter $\alpha(\tau)$. At least conceptually, this can be done in a straightforward manner: (1) Obtain an estimator $\widehat{\beta}(\tau,\alpha(\tau))$,[10] given the hypothesis of $\alpha(\tau)$ from the moment condition: $m(\alpha(\tau),\beta) = 0$ when $\beta = \beta(\tau,\alpha(\tau)) = p\lim\widehat{\beta}(\tau,\alpha(\tau))$.[11] (2) Plug in $\widehat{\beta}(\tau,\alpha(\tau))$ to the joint test statistic. That is, find $K_n^*(\alpha(\tau),\widehat{\beta}(\tau,\alpha(\tau)))$. (3) Adjust the degrees of freedom. As long as $\frac{\partial m(\alpha(\tau),\beta)}{\partial\beta'}\mid_{\beta=\beta(\tau,\alpha(\tau))}$ has a full column rank, the resulting statistic, $K_n^*(\alpha(\tau),\widehat{\beta}(\tau,\alpha(\tau)))$, will be robust to weak values of $\frac{\partial m(\alpha,\beta)}{\partial\alpha}\mid_{\alpha=\alpha(\tau),\beta=\beta(\tau,\alpha(\tau))}$ (see e.g., Kleibergen 2002, 2004, 2005). The difficulty is that it is computationally infeasible because finding $\widehat{\beta}(\tau,\alpha(\tau))$ from the given moment condition requires a grid search over a potentially high dimensional space.

To overcome this computational difficulty, it is noted that finding an estimator $\widehat{\beta}(\tau,\alpha(\tau))$ *using* $m_1(\alpha(\tau),\beta) = 0$ *only* does not require the grid search because it is simply a conventional quantile regression problem that can be easily solved by linear programming (see Koenker and Basset (1978) and Buchinsky (1998), among others). Therefore, the following strategy is taken. First, define $\beta(\tau,\alpha)$ that solves $m_1(\alpha,\beta) = 0$ for each $\alpha$ and consider the concentrated moment condition, $m_c(\alpha) \equiv m_2(\alpha,\beta(\tau,\alpha)) = E(W_i(1\{\overline{Y}_i \leq D_i\alpha + X_i'\beta(\tau,\alpha)\} - \tau)) = 0$ when $\alpha = \alpha(\tau)$. The *AR* statistic testing $\alpha(\tau)$ can be easily obtained by this two-step strategy (Honore and Hu (2004)). However, the power issue arises due to wasted degrees of freedom; therefore, this paper pursues the orthogonal decomposition of this two-step *AR* statistic. For this purpose, the Jacobian of the concentrated moment condition is considered. First, the following assumptions are made.

---

[9] $AR_n^*(\alpha(\tau),\beta(\tau)) = K_n^*(\alpha(\tau),\beta(\tau)) + J_n^*(\alpha(\tau),\beta(\tau))$,
where $J_n^*(\alpha(\tau),\beta(\tau)) \equiv (\sqrt{n}\widehat{m})'\widetilde{V}_{11}^{-\frac{1}{2}}(I - P_{\widetilde{V}_{11}^{-\frac{1}{2}}\widehat{J}})\widetilde{V}_{11}^{-\frac{1}{2}}(\sqrt{n}\widehat{m})$.

[10] $\widehat{\beta}(\tau,\alpha(\tau))$ emphasizes that it is obtained under the null hypothesis of $\alpha(\tau)$.

[11] Of course, under the null of $\alpha(\tau)$, $\beta(\tau,\alpha(\tau)) = \beta(\tau)$.

**Assumption 3** *$\epsilon \equiv \overline{Y} - D\alpha(\tau) - X'\beta(\tau, \alpha(\tau))$ has a well defined density conditional on $D, Z, X$, $f_\epsilon(s \mid D, Z, X)$ and it is bounded away from 0 at $s = 0$, where $\beta(\tau, \alpha)$ solves $m_1(\alpha, \beta) = 0$ for each $\alpha$.*

**Assumption 4** *$E(XX'f_\epsilon(0 \mid D, Z, X))$ is non-singular.*

Assumption 4 requires that $\beta(\tau, \alpha)$ is well identified from the first set of moment conditions, $m_1(\alpha, \beta(\tau, \alpha)) = 0$ for each $\alpha$.

**Lemma 2** *Under assumptions 3 and 4, we have $\Gamma \equiv \Gamma(\alpha(\tau), \beta(\tau, \alpha(\tau))) \equiv \frac{\partial m_c(\alpha)}{\partial \alpha} \mid_{\alpha=\alpha(\tau)}$ $= E(QDf_\epsilon(0 \mid D, Z, X))$, where $Q \equiv W - AX$ with $A \equiv E(WX'f_\epsilon(0 \mid D, Z, X))E(XX'f_\epsilon(0 \mid D, Z, X))^{-1}$.*

**Proof.** See appendix. ■

Note that when $E(QDf_\epsilon(0 \mid D, Z, X)) = 0$, $\alpha(\tau)$ is not identified. The following sections discuss a test of $\alpha(\tau)$ that is robust to weak values of $E(QDf_\epsilon(0 \mid D, Z, X))$. It will be first assumed that $\beta(\tau, \alpha(\tau))$ is known, and this assumption will be removed in section 5. Note also that $Q$ is the transformed instrument which will be different from $W$ unless $A = 0$. Throughout the discussion, it will also be assumed that the transformed instrument $Q$ is known. Then, using the estimated $Q$ will be briefly discussed at the end of section 5 (see also appendix).

# 4 Testing $\alpha(\tau)$ with a Known $\beta(\tau, \alpha(\tau))$

This section considers the test of $\alpha(\tau)$ with the assumption that $\beta(\tau, \alpha(\tau))$ is known. It will be shown that the proposed test statistic is robust to weak values of $\Gamma \equiv E(QDf_\epsilon(0 \mid D, Z, X))$. As mentioned before, directly smoothing the objective function with a cdf type kernel is not pursued because it raises a bias issue from the first step of defining the $AR$ statistic. Nevertheless, it is useful to estimate $\Gamma = E(QDf_\epsilon(0 \mid D, Z, X))$ by considering its derivative. Hence, a hybrid approach is taken where the indicator is smoothed only to deal with the derivative part while the indicator in the moment condition itself is left unsmoothed. In order to achieve this, the following assumptions are made. For the sake of notational simplicity, let $Y = \overline{Y} - X'\beta(\tau, \alpha(\tau))$. The data are assumed to be *iid*.

**Assumption 5** *We have a kernel* $k(v)$ *such that* $\sup | k(v) | < \infty$, $\int | k(v) | dv < \infty$, $\int k(v)dv = 1$, $\int k(v)^2 dv < \infty$, *and* $| v || k(v) | \to 0$ *as* $| v | \to 0$.

**Assumption 6** $f_\epsilon(s \mid D, Z, X)$ *is twice differentiable at* $s = 0$.

These are the standard assumptions for the kernel of density estimation. The density of $N(0,1)$ is common for $k(v)$.

**Lemma 3** *Let* $\widehat{\Gamma}_{n,h} = \frac{1}{nh}\sum_{i=1}^n Q_i D_i k(\frac{-\epsilon_i}{h})$, *where* $h$ *is a bandwidth choice. Then,* $\widehat{\Gamma}_{n,h} - \Gamma \xrightarrow{p} 0$ *as* $h \downarrow 0$ *and* $nh \to \infty$. *In particular, for any sequence* $h_n > 0$ *with* $h_n \to 0, nh_n \to \infty$, *we have* $\widehat{\Gamma}_{n,h_n} \xrightarrow{p} \Gamma$ *as* $n \to \infty$.

**Proof.** See appendix. ■

This lemma suggests using $\widehat{\Gamma}_{n,h_n}$ for the robust statistic. For this purpose, the joint distribution of $\sqrt{n}\widehat{m}_c \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n W_i(1\{Y_i \le D_i\alpha(\tau)\} - \tau)$ and $\sqrt{nh_n}(\widehat{\Gamma}_{n,h_n} - \Gamma)$ is considered in the following assumption. Note that $W_i$ and $Q_i = W_i - AX_i$ are being used for $\widehat{m}_c$ and $\widehat{\Gamma}_{n,h_n}$, respectively.

**Assumption 7** *For any* $h > 0$, *suppose*

$$Var(\begin{bmatrix} \sqrt{n}\widehat{m}_c \\ \sqrt{nh}(\widehat{\Gamma}_{n,h} - E(\widehat{\Gamma}_{n,h})) \end{bmatrix}) = \begin{bmatrix} V & V_{12,h} \\ V_{21,h} & V_{22,h} \end{bmatrix},$$ *where* $V$ *is positive definite.*

*Then, for any sequence* $h_n > 0$ *with* $h_n \to 0$, $nh_n \to \infty$, *we have*

$$\begin{bmatrix} V & V_{12,h_n} \\ V_{21,h_n} & V_{22,h_n} \end{bmatrix}^{-\frac{1}{2}} \begin{bmatrix} \sqrt{n}\widehat{m}_c \\ \sqrt{nh_n}(\widehat{\Gamma}_{n,h_n} - E(\widehat{\Gamma}_{n,h_n})) \end{bmatrix} \xrightarrow{d} N(0, I) \text{ as } n \to \infty.$$

Assumption 7 is a central limit theorem with a sequential $h_n$ and it is satisfied under mild conditions such as $\int k(v)^{2+\delta}dv < \infty$ for any $\delta > 0$. Note also that the dependency of the variance on the null hypothesis is suppressed for the sake of notational simplicity.

**Lemma 4** *As* $h \downarrow 0$, *we have* $V_{21,h} = V'_{12,h} \to 0$ *and* $V_{22,h} \to V_{22}$. *Also, we have* $\sqrt{nh}(E(\widehat{\Gamma}_{n,h}) - \Gamma) \to B = \begin{cases} c_1 E(Q_i D_i(-f'_\epsilon(0 \mid D_i, Z_i, X_i)) \int vk(v)dv & \text{if } \sqrt{nh}h \to c_1 \\ c_2 \frac{1}{2} E(Q_i D_i f''_\epsilon(0 \mid D_i, Z_i, X_i)) \int v^2 k(v)dv & \text{if } \sqrt{nh}h^2 \to c_2 \text{ and } \int vk(v)dv = 0 \end{cases}.$

*In particular, for any sequence* $h_n > 0$ *with* $h_n \to 0, nh_n \to \infty$, *and* $\sqrt{nh_n}h_n \to c_1$ *(or* $\sqrt{nh_n}h_n^2 \to c_2$

*and* $\int vk(v)dv = 0$*), we have*

$$\begin{bmatrix} \sqrt{n}\widehat{m}_c \\ \sqrt{nh_n}(\widehat{\Gamma}_{n,h_n} - \Gamma) \end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix} 0 \\ B \end{bmatrix}, \begin{bmatrix} V & 0 \\ 0 & V_{22} \end{bmatrix}\right) \text{ as } n \to \infty.$$

**Proof.** See appendix. ■

Notice that the asymptotic independence is due to the difference in the convergence rates of the top and bottom parts. If the true conditional density $f_\epsilon(0 \mid D, Z, X)$ were known, $\Gamma$ could be estimated with $\sqrt{n}$ rate. However, because the true conditional density is not known, it is not possible to do as good a job and the convergence rate becomes slower than $\sqrt{n}$. The slower convergence rate of the bottom makes it asymptotically independent of the top part.

**Theorem 2** *Consider a sequence* $h_n > 0$ *with* $h_n \to 0, nh_n \to \infty$*, and* $\sqrt{nh_n}h_n \to c_1$ *(or* $\sqrt{nh_n}h_n^2 \to c_2$ *and* $\int vk(v)dv = 0$*) and let* $S_n \equiv (\sqrt{nh_n}\widehat{\Gamma}_{n,h_n})'V^{-1}(\sqrt{n}\widehat{m}_c)$*.*
*Define* $K_n \equiv S_n'((\sqrt{nh_n}\widehat{\Gamma}_{n,h_n})'V^{-1}(\sqrt{nh_n}\widehat{\Gamma}_{n,h_n}))^{-1}S_n$*.*
*Then,* $K_n \xrightarrow{d} \chi^2(1)$ *as* $n \to \infty$ *when* $\Gamma$ *is (i) 0, (ii)* $O(\frac{1}{\sqrt{n}})$*, (iii)*$O(\frac{1}{\sqrt{nh_n}})$ *or (iv) of full rank.*

**Proof.** See appendix. ■

Since $K_n$ can be written as $(\sqrt{n}\widehat{m}_c)'V^{-\frac{1}{2}}P_{V^{-\frac{1}{2}}\widehat{\Gamma}_{n,h_n}}V^{-\frac{1}{2}}(\sqrt{n}\widehat{m}_c)$, the orthogonal decomposition of the $AR$ statistic can be obtained as before: [12]

$$\begin{aligned} AR_n &\equiv (\sqrt{n}\widehat{m}_c)'V^{-1}(\sqrt{n}\widehat{m}_c) \xrightarrow{d} \chi^2(\dim(W)) \\ &= K_n + J_n \end{aligned}$$

where $J_n \equiv (\sqrt{n}\widehat{m}_c)'V^{-\frac{1}{2}}(I - P_{V^{-\frac{1}{2}}\widehat{\Gamma}_{n,h_n}})V^{-\frac{1}{2}}(\sqrt{n}\widehat{m}_c)$. $K_n$ and $J_n$ are asymptotically independent by construction.

Some remarks are necessary. First, while $K_n$ is not a score statistic from the GMM objective function whose derivative is almost everywhere 0, it is based on the score of the limit function. It implies that $K_n$ is likely to show some spurious power declines over the points where the score function is actually 0. This is a well known phenomenon and it can be overcome by using the suggested orthogonal decomposition. The next section will explain how to use these orthogonal components for conducting inferences; that section also develops the feasible versions of these statistics. Second, comparing it with the Wald statistic, the distribution of $\sqrt{n}\widehat{m}_c$ is directly used

[12] $P_X$ is the projection onto the column space of $X$.

without trying to connect it with $\sqrt{n}(\widehat{\alpha}(\tau) - \alpha(\tau))$. In this sense, $K_n$ shares the spirit of the $AR$ statistic. The difference from the $AR$ statistic is that $K_n$ does not use the whole $\sqrt{n}\widehat{m}_c$ but only uses its projection onto a certain space by which the degrees of freedom of the $AR$ statistic is reduced. Since the $AR$ statistic is the joint test of the location of $\alpha(\tau)$ and the overidentifying restrictions, it is hard to interpret inferences based on the $AR$ statistic. This decomposition provides a better means of interpretation. Lastly, it is pointed out that these statistics are infeasible because $\beta(\tau, \alpha(\tau))$ and $Q_i = W_i - AX_i$ are not known.

## 5 Testing $\alpha(\tau)$ with an Unknown $\beta(\tau, \alpha(\tau))$

It has been assumed so far that the true $\beta(\tau, \alpha(\tau))$ is known because once the location of $\alpha(\tau)$ is hypothesized, $\beta(\tau, \alpha(\tau))$ can be easily estimated by conventional quantile regression (Koenker and Basset (1978)) without any computational difficulty. In this section, relaxing the assumption of the known $\beta(\tau, \alpha(\tau))$ is explicitly considered and the feasible counterparts of $K_n$, $J_n$ will be proposed.

First, let $\Upsilon_{n,h} \equiv \begin{bmatrix} \sqrt{n}I & 0 \\ 0 & \sqrt{nh}I \end{bmatrix}$ and define

$$\begin{bmatrix} \widehat{m}_c(\alpha(\tau), \beta) \\ \widehat{\Gamma}_{n,h}(\alpha(\tau), \beta) \end{bmatrix} \equiv \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} W_i(1\{\overline{Y}_i \leq D_i\alpha(\tau) + X_i'\beta\} - \tau) \\ \frac{1}{nh}\sum_{i=1}^{n} Q_i D_i k(\frac{D_i\alpha(\tau) + X_i'\beta - \overline{Y}_i}{h}) \end{bmatrix} \tag{5.1}$$

For the sake of convenience of discussion, some definitions are made.

**Definition 1** *For any $h > 0$, define* $G_{n,h}(\alpha(\tau), \beta) \equiv \begin{bmatrix} \widehat{m}_c(\alpha(\tau), \beta) \\ \widehat{\Gamma}_{n,h}(\alpha(\tau), \beta) - E(\widehat{\Gamma}_{n,h}(\alpha(\tau), \beta(\tau, \alpha(\tau)))) \end{bmatrix}$ *and* $G_h(\alpha(\tau), \beta) \equiv E(G_{n,h}(\alpha(\tau), \beta)) \equiv \begin{bmatrix} m_c(\alpha(\tau), \beta) \\ E(\widehat{\Gamma}_{n,h}(\alpha(\tau), \beta)) - E(\widehat{\Gamma}_{n,h}(\alpha(\tau), \beta(\tau, \alpha(\tau)))) \end{bmatrix}$.

*Note* $G_h(\alpha(\tau), \beta(\tau, \alpha(\tau))) = 0$.

The sub-index $h$ in $G_h(\alpha(\tau), \beta)$ indicates that its bottom part depends on $h$. Subtracting $E(\widehat{\Gamma}_{n,h}(\alpha(\tau), \beta(\tau, \alpha(\tau))))$ is a normalization for convenience.

Then, the following assumptions are made.

**Assumption 8** *There is an estimator $\widehat{\beta}(\tau, \alpha(\tau))$ that is asymptotically linear with respect to some influence function $\Xi_i$. That is, $\sqrt{n}(\widehat{\beta}(\tau, \alpha(\tau)) - \beta(\tau, \alpha(\tau))) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \Xi_i + o_p(1)$ where $\Xi_i$ are iid with mean zero. $\widehat{\beta}(\tau, \alpha(\tau))$ can be obtained by conventional quantile regression of $\overline{Y} - D\alpha(\tau)$ on $X$.*

**Assumption 9** *$G_h(\alpha(\tau),\beta) \to G(\alpha(\tau),\beta) \equiv \begin{bmatrix} m_c(\alpha(\tau),\beta) \\ \Gamma(\alpha(\tau),\beta) - \Gamma(\alpha(\tau),\beta(\tau,\alpha(\tau))) \end{bmatrix}$ uniformly, at least locally around $\beta(\tau,\alpha(\tau))$ as $h \downarrow 0$.*

**Assumption 10** *$G(\alpha(\tau),\beta)$ can be linearized with respect to $\beta$ around $\beta(\tau,\alpha(\tau))$.
That is, $G(\alpha(\tau),\beta) = H(\beta - \beta(\tau,\alpha(\tau))) + o(\| \beta - \beta(\tau,\alpha(\tau)) \|)$ for some $H$.
We will write $H = \begin{bmatrix} H_1' & H_2' \end{bmatrix}'$ where $H_1, H_2$ are from the top and the bottom parts of $G$.*

**Assumption 11** *For any $h > 0$, suppose
$Var(\Upsilon_{n,h}G_{n,h}(\alpha(\tau),\beta(\tau,\alpha(\tau))) + \begin{bmatrix} H_1\frac{1}{\sqrt{n}}\sum_{i=1}^n \Xi_i \\ 0 \end{bmatrix}) = \begin{bmatrix} V^* & V_{12,h}^* \\ V_{21,h}^* & V_{22,h}^* \end{bmatrix}$, where $V^*$ is positive definite.
Then, for any sequence $h_n > 0$ with $h_n \to 0, nh_n \to \infty$, we have*

$$\begin{bmatrix} V^* & V_{12,h_n}^* \\ V_{21,h_n}^* & V_{22,h_n}^* \end{bmatrix}^{-\frac{1}{2}} \{\Upsilon_{n,h_n}G_{n,h_n}(\alpha(\tau),\beta(\tau,\alpha(\tau))) + \begin{bmatrix} H_1\frac{1}{\sqrt{n}}\sum_{i=1}^n \Xi_i \\ 0 \end{bmatrix})\} \xrightarrow{d} N(0,I) \text{ as } n \to \infty.$$

**Assumption 12** *For any $\delta_n \downarrow 0$ and for any sequence $h_n > 0$ with $h_n \to 0, nh_n \to \infty$, we have*

$$\sup_{\|\beta-\beta(\tau,\alpha(\tau))\|<\delta_n} \| \Upsilon_{n,h_n}(G_{n,h_n}(\alpha(\tau),\beta) - G_{h_n}(\alpha(\tau),\beta)) - \Upsilon_{n,h_n}(G_{n,h_n}(\alpha(\tau),\beta(\tau,\alpha(\tau))) - G_{h_n}(\alpha(\tau),\beta(\tau,\alpha(\tau))) \| = o_p(1) \quad \text{as } n \to \infty.$$

**Assumption 13** *For any seqeunce $h_n > 0$ with $h_n \to 0, nh_n \to \infty$ and $\sqrt{nh_n}h_n \to c_1$ (or $\sqrt{nh_n}h_n^2 \to c_2$ if $\int vk(v)dv = 0$), $B_{n,h_n}(\alpha(\tau),\beta) \equiv \sqrt{nh_n}(E(\widehat{\Gamma}_{n,h_n}(\alpha(\tau),\beta)) - \Gamma(\alpha(\tau),\beta)) \to B(\alpha(\tau),\beta)$ uniformly in $\beta$, at least locally around $\beta(\tau,\alpha(\tau))$ as $n \to \infty$.
We also assume that $B(\alpha(\tau),\beta), \Gamma(\alpha(\tau),\beta)$ are continuous at $\beta = \beta(\tau,\alpha(\tau))$.*

Assumption 12 is a stochastic equicontinuity with a Lindberg type condition (see e.g., van der Vaart and Wellner (1996, theorem 2.11.1) for primitive conditions). It is a restriction on the class of functions involved in $G_{n,h_n}(\alpha(\tau),\beta)$. Since the indicator function and the family of parametric densities satisfying the Lipschitz condition are typically allowed, this assumption holds with various choices of kernels.

For assumptions 9 and 13, the pointwise convergences of $G_{h_n}(\alpha(\tau),\beta)$ and $B_{n,h_n}(\alpha(\tau),\beta)$ can be verified by direct computation due to lemma A in the appendix. The local uniformity is satisfied when $G_{h_n}(\alpha(\tau),\beta)$ and $B_{n,h_n}(\alpha(\tau),\beta)$ are equicontinuous.[13]

[13] $g_n(x)$ is said to converge to $g(x)$ locally uniformly around $x_0$ when for any sequence $\delta_n \downarrow 0$, $\sup_{\|x-x_0\|\leq\delta_n} \|$

**Lemma 5** *Under these assumptions, for any sequence $h_n > 0$ with $h_n \to 0, nh_n \to \infty$, we have* $\Upsilon_{n,h_n}(G_{n,h_n}(\alpha(\tau), \widehat{\beta}(\tau, \alpha(\tau))) = \Upsilon_{n,h_n} G_{n,h_n}(\alpha(\tau), \beta(\tau, \alpha(\tau))) + \begin{bmatrix} H_1\sqrt{n}(\widehat{\beta}(\tau, \alpha(\tau)) - \beta(\tau, \alpha(\tau))) \\ 0 \end{bmatrix}$ $+o_p(1)$.

**Proof.** See appendix. ■

**Lemma 6** *As $h \downarrow 0$, we have $V^*_{21,h} = V^{*\prime}_{12,h} \to 0$ and $V^*_{22,h} \to V^*_{22}$.*
*In particular, for any sequence $h_n > 0$ with $h_n \to 0, nh_n \to \infty$, and $\sqrt{nh_n}h_n \to c_1$ (or $\sqrt{nh_n}h_n^2 \to c_2$ and $\int vk(v)dv = 0$), we have*

$$\begin{bmatrix} \sqrt{n}\widehat{m}_c(\alpha(\tau), \widehat{\beta}(\tau, \alpha(\tau))) \\ \sqrt{nh_n}(\widehat{\Gamma}_{n,h_n}(\alpha(\tau), \widehat{\beta}(\tau, \alpha(\tau))) - \Gamma(\alpha(\tau), \beta(\tau, \alpha(\tau)))) \end{bmatrix} \xrightarrow{d} N(\begin{bmatrix} 0 \\ B(\alpha(\tau), \beta(\tau, \alpha(\tau))) \end{bmatrix}, \begin{bmatrix} V^* & 0 \\ 0 & V^*_{22} \end{bmatrix})$$

*as $n \to \infty$.*

**Proof.** See appendix. ■

Lemma 6 shows that the robust statistics can be obtained by using $\widehat{\beta}(\tau, \alpha(\tau))$ instead of $\beta(\tau, \alpha(\tau))$. However, unless $H_1 = \frac{\partial m_c(\alpha(\tau),\beta)}{\partial \beta'} |_{\beta=\beta(\tau,\alpha(\tau))} = 0$, the variance matrix is different from that of the previous section and the variance effect of $\widehat{\beta}(\tau, \alpha(\tau))$ should be taken into account as is standard in the two-step estimator theory (see Newey and McFadden (1994)). Generally, $V^*$ can be estimated by either kernel smoothing or bootstrap. Bootstrap is, however, not computationally attractive because $V^*$ is the continuous updating variance that depends on $\alpha(\tau)$. To construct a confidence set, the bootstrap would have to be repeated for every hypothesized $\alpha(\tau)$. For the kernel smoothing estimation of $V^*$, see the appendix.

**Theorem 3** *For any sequence $h_n > 0$ with $h_n \to 0, nh_n \to \infty$, and $\sqrt{nh_n}h_n \to c_1$ (or $\sqrt{nh_n}h_n^2 \to c_2$ if $\int vk(v)dv = 0$), let $\widehat{S}_n \equiv (\sqrt{nh_n}\widehat{\Gamma}_{n,h_n}(\alpha(\tau), \widehat{\beta}(\tau, \alpha(\tau))))' V^{*-1}(\sqrt{n}\widehat{m}_c(\alpha(\tau), \widehat{\beta}(\tau, \alpha(\tau))))$.*
*Define $\widehat{K}_n \equiv \widehat{S}_n'((\sqrt{nh_n}\widehat{\Gamma}_{n,h_n}(\alpha(\tau), \widehat{\beta}(\tau, \alpha(\tau))))' V^{*-1}(\sqrt{nh_n}\widehat{\Gamma}_{n,h_n}(\alpha(\tau), \widehat{\beta}(\tau, \alpha(\tau)))))^{-1}\widehat{S}_n$.*
*Then, $\widehat{K}_n \xrightarrow{d} \chi^2(1)$ as $n \to \infty$ when $\Gamma(\alpha(\tau), \beta(\tau, \alpha(\tau)))$ is (i) 0, (ii) $O(\frac{1}{\sqrt{n}})$, (iii)$O(\frac{1}{\sqrt{nh_n}})$ or (iv) of full rank.*

---

$g_n(x) - g(x) \| \to 0$. Notice that it clearly holds when $g_n(x)$ is equicontinuous, $g(x)$ is continuous at $x_0$, and $g_n(x) \to g(x)$ at each $x$ because we then have $g_n(x_n) \to g(x_0)$ for any $x_n$ converging to $x_0$.

Instead of providing primitive conditions for the equicontinuity, only the necessary high-level assumptions are stated.

**Proof.** Follow the proof of theorem 2, using lemma 6. ■

$\widehat{K}_n$ is the counterpart of $K_n$ in using $\widehat{\beta}(\tau,\alpha(\tau))$. Strictly speaking, $\widehat{K}_n$ is still infeasible because the transformed instruments $Q_i = W_i - AX_i$ are not known, where $A = E(WX'f_\epsilon(0 \mid D,Z,X))E(XX'f_\epsilon(0 \mid D,Z,X))^{-1}$. However, notice that $A$ can be consistently estimated by $\widehat{A}_{n,h_n} = (\frac{1}{nh_n}\sum_{i=1}^n W_i X_i' k(\frac{D_i\alpha(\tau)+X_i'\widehat{\beta}(\tau,\alpha(\tau))-\overline{Y}_i}{h_n}))(\frac{1}{nh_n}\sum_{i=1}^n X_i X_i' k(\frac{D_i\alpha(\tau)+X_i'\widehat{\beta}(\tau,\alpha(\tau))-\overline{Y}_i}{h_n}))^{-1}$. In fact, the convergence rate of $\widehat{A}_{n,h_n}$ is given by $\sqrt{nh_n}$ under mild conditions; therefore, using $\widehat{A}_{n,h_n}$ instead of the true $A$ will only change the asymptotic variance $V_{22}^*$, except for some additional bias to $B(\alpha(\tau),\beta(\tau))$ in lemma 6. Since the bias and the variance of the bottom part do not matter for the robust statistics, $\widehat{A}_{n,h_n}$ can be used in place of the true $A$ without changing the main results of theorem 3. The assumptions needed for this argument are exactly parallel to those in this section and thus are not stated here. See appendix.

Note also that the orthogonal decomposition of the $AR$ statistic is similarly obtained: $\widehat{J}_n \equiv \widehat{AR}_n - \widehat{K}_n$, where $\widehat{K}_n, \widehat{J}_n$ are asymptotically independent by construction. Therefore, the following procedures are proposed in order to test $\alpha(\tau)$. Let $crt_{\alpha,k}$ be the $\alpha$% critical value from $\chi^2(k)$.

(1) Two-step $AR$ test: Joint test of the location $\alpha(\tau)$ and the over-identifying restrictions.

- Step 1: Obtain $\widehat{\beta}(\tau,\alpha(\tau))$ from quantile regression of $\overline{Y}_i - D_i\alpha(\tau)$ on $X_i$.
- Step 2: Obtain $\widehat{AR}_n \equiv (\sqrt{n}\widehat{m}_c(\alpha(\tau),\widehat{\beta}(\tau,\alpha(\tau)))'V^{*-1}(\sqrt{n}\widehat{m}_c(\alpha(\tau),\widehat{\beta}(\tau,\alpha(\tau))))$.
  - Reject $\alpha(\tau)$ when $\widehat{AR}_n > crt_{\alpha,\dim(W)}$.

(2) Three-step decomposition of the $AR$ statistic.

- Step 1: Obtain $\widehat{\beta}(\tau,\alpha(\tau))$ from quantile regression of $\overline{Y}_i - D_i\alpha(\tau)$ on $X_i$.
- Step 2: Obtain $\widehat{A}_{n,h_n} = (\frac{1}{nh_n}\sum_{i=1}^n W_i X_i' k(\frac{D_i\alpha(\tau)+X_i'\widehat{\beta}(\tau,\alpha(\tau))-\overline{Y}_i}{h_n}))(\frac{1}{nh_n}\sum_{i=1}^n X_i X_i' k(\frac{D_i\alpha(\tau)+X_i'\widehat{\beta}(\tau,\alpha(\tau))-\overline{Y}_i}{h_n}))^{-1}$.
- Step 3: Obtain $\widehat{K}_n$ by using $\widehat{Q}_i \equiv W_i - \widehat{A}_{n,h_n}X_i$.
  - Orthogonal decomposition of $\widehat{AR}_n$: $\widehat{J}_n \equiv \widehat{AR}_n - \widehat{K}_n$
  - Modified $K$ test: Reject $\alpha(\tau)$ when $\widehat{K}_n > crt_{\alpha,1}$.

– Modified $K$-$J$ combination: Reject $\alpha(\tau)$ when $\widehat{K}_n > crt_{\alpha 1,1}$ or $\widehat{J}_n > crt_{\alpha 2, \dim(W)-1}$, where $\alpha 1 + \alpha 2 = \alpha$.[14]

The decomposition of the $AR$ statistic is computationally attractive since only $V^*$ is used for each orthogonal component. Kleibergen (2005) also suggested the GMM counterpart to the conditional likelihood ratio statistic of Moreira (2003). Kleibergen considered conditioning on a rank statistic testing the reduced rank of $\Gamma$. In principle, this approach can be extened, but such extension is not considered here.[15] For the estimation of $V^*$, either kernel smoothing or bootstrap can be used. Bootstrap is, however, not computationally attractive because $V^*$ is the continuous updating variance that depends on $\alpha(\tau)$. To obtain a confidence set, the bootstrap will have to be repeated for every hypothesized $\alpha(\tau)$. For the kernel smoothing estimation of $V^*$, see appendix.

# 6 Simulations

For simulation purposes, the simplest location model is used. The experiments use the following design.

$$\begin{aligned} \overline{Y}_i &= \beta(U_i) + D_i \alpha_0 \\ D_i &= \gamma_1 + Z_i' \gamma_2 + V_i \end{aligned} \tag{6.1}$$

where $\beta(s) = \Phi^{-1}(s), U_i \sim uniform(0,1)$. This is a location model in the sense that $\alpha(\tau) = \alpha_0$ for all $\tau \in (0,1)$. It is also assumed that $Z_i = [Z_{2i}, Z_{3i}, ...., Z_{9i}]' \sim iid\ N(0, I_8)$,[16] $[\beta(U_i), V_i]' \sim iid\ N(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix})$, and that $Z_i$ and $[\beta(U), V]'$ are independent. Note that $\rho \neq 0$ implies the endogeneity of $D_i$. The parameter of interest is the $\tau^{th}$ quantile effect of the exogenous change of $D_i$. That is, the goal is testing $H_0 : \alpha(\tau) = \alpha_H$ for a given $\tau$.[17] If it is known a priori that $\alpha(\tau) = \alpha_0$ for all $\tau$, then this is a standard simultaneous equation model and there are many other methods

[14] Note that $\Pr(\widehat{K}_n > crt_{\alpha 1,1}$ or $\widehat{J}_n > crt_{\alpha 2, \dim(W)-1}) = \alpha 1 + \alpha 2 - \alpha 1 * \alpha 2 \approx \alpha 1 + \alpha 2 = \alpha$ by the independence of $\widehat{K}_n$ and $\widehat{J}_n$.

[15] This will require estimating the variance of the bottom part, which is affected by the estimation of $\widehat{A}_{n,h_n}$. See appendix.

[16] $I_k$ is a $k \times k$ identity matrix.

[17] We distinguish the true value $\alpha_0$ from the hypothesized value $\alpha_H$.

to estimate and test $\alpha_0$, such as 2SLS and LIML. Nevertheless, this design is used for the following reasons. First, the instrumental quantile method is sometimes preferred even in this simple location model because of its robustness to the outliers (Chen and Portnoy (1996), Honore and Hu (2004)). Second, the possibility of quantile effect heterogeneity should be allowed for, and thus assuming a priori that $\alpha(\tau) = \alpha_0$ for all $\tau$ is not desirable. Third, the purpose of the experiments is to see the power properties of the developed methods under the various values of the Jacobian of the limit function, and thus a sufficiently simple model with the analytically tracktable Jacobian is preferred.

For a given $\tau \in (0,1)$, let $m(\alpha,\beta) \equiv \begin{bmatrix} m_1(\alpha,\beta) & m_2(\alpha,\beta)' \end{bmatrix}' = E(Q_{Xi}(1\{\overline{Y}_i \leq D_i\alpha + \beta\} - \tau))$, where $Q_{Xi} = \begin{bmatrix} 1 & Z_i' \end{bmatrix}'$. Define $\beta(\tau,\alpha)$ that solves $m_1(\alpha,\beta) = 0$ for each $\alpha$, then consider the concentrated moment condition: $m_c(\alpha) = m_2(\alpha, \beta(\tau,\alpha)) = 0$ when $\alpha = \alpha(\tau) = \alpha_0$ for any given $\tau$.

Note that the designed setup is simple enough to analytically compute the Jacobian evaluated at $\alpha = \alpha_0$.

$$\frac{\partial m_c}{\partial \alpha} |_{\alpha=\alpha_0} = \Gamma = (2\pi)^{-\frac{1}{2}} \exp(-\frac{\beta(\tau,\alpha_0)^2}{2}) Var(Z_i)\gamma_2 \tag{6.2}$$

For the derivation, see appendix.

The full rank assumption of $\Gamma$ hinges on several points. The variation of $Z_i$ should be sufficiently rich and $\gamma_2$ should be non-zero. Given the variation of $Z_i$, when $\gamma_2$ is too close to 0, the identification becomes weak. Another interesting feature is that the local identification is also affected by $\tau$. Recall that $\beta(\tau,\alpha_0)$ is the $\tau^{th}$ quantile of $\overline{Y}_i - D_i\alpha_0 = \beta(U_i)$. Hence, it is easy to see that $\beta(\tau,\alpha_0) = \beta(\tau) = \Phi^{-1}(\tau)$. Since $\exp(-\frac{(\Phi^{-1}(\tau))^2}{2})$ is maximized at $\tau = 0.5$, the median effect, $\alpha(0.5)$, is best identified. When the given instruments are weak, the identification of the tail quantiles will suffer more seriously than that of the median effect.

The fact that the identification depends on $\tau$ is interesting. Although the median is better identified than the tails in this simple setup, this is purely due to the artificial design. In practice, it is never clear for what quantiles the given instruments will be most informative.

In the power experiments, the sample size $n = 500$ was used, and 1000 repetitions were made for Monte Carlo. $\rho$ was set to be 0.8 so that $D$ was highly endogenous. $\gamma_1$ and $\gamma_2$ were set to be 1 and $w * \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}'$, respectively, where $w \in \{0,\ 0.02, 0.05, 0.10, 0.20, 1.00\}$. When $w = 0$, there is no identification. When $w = 0.02$, 0.05, the model is weakly identified and $w = 1.00$ is the strong identification case. $\tau = 0.25$, 0.50, 0.75 were considered.

Figures 1 through 3 illustrate the power curves of the various test statistics. $H_0 : \alpha(\tau) = 1$ was considered as the various values of $\alpha_0$ were tried. $h_n = n^{-\frac{1}{5}}$ was used as the bandwidth choice.[18] For expository purposes, the power curves of the Wald statistic from the smoothed GMM[19] and the Wald statistic from the dual inference method[20] proposed by Chernozhukov and Hansen (2004) were also computed. The dual method, $AR$, $K$, $J$, and $K$-$J$ combination give the approximately correct size regardless of the quality of identification. In case of the complete lack of identification, none of these statistics have any spurious power. The power declines of the $K$ statistic can be explained by the shape of the limit function since it is based on the score of the limit function. Note that the power declines of the $K$ statistic could be easily fixed by combining it with the $J$ statistic which is the other orthogonal component of the $AR$ statistic. It is also worth noting that the power curve of the dual inference method is similar to that of the $AR$ statistic. This is not surprising because both of them test the location and the overidentifying restrictions jointly.

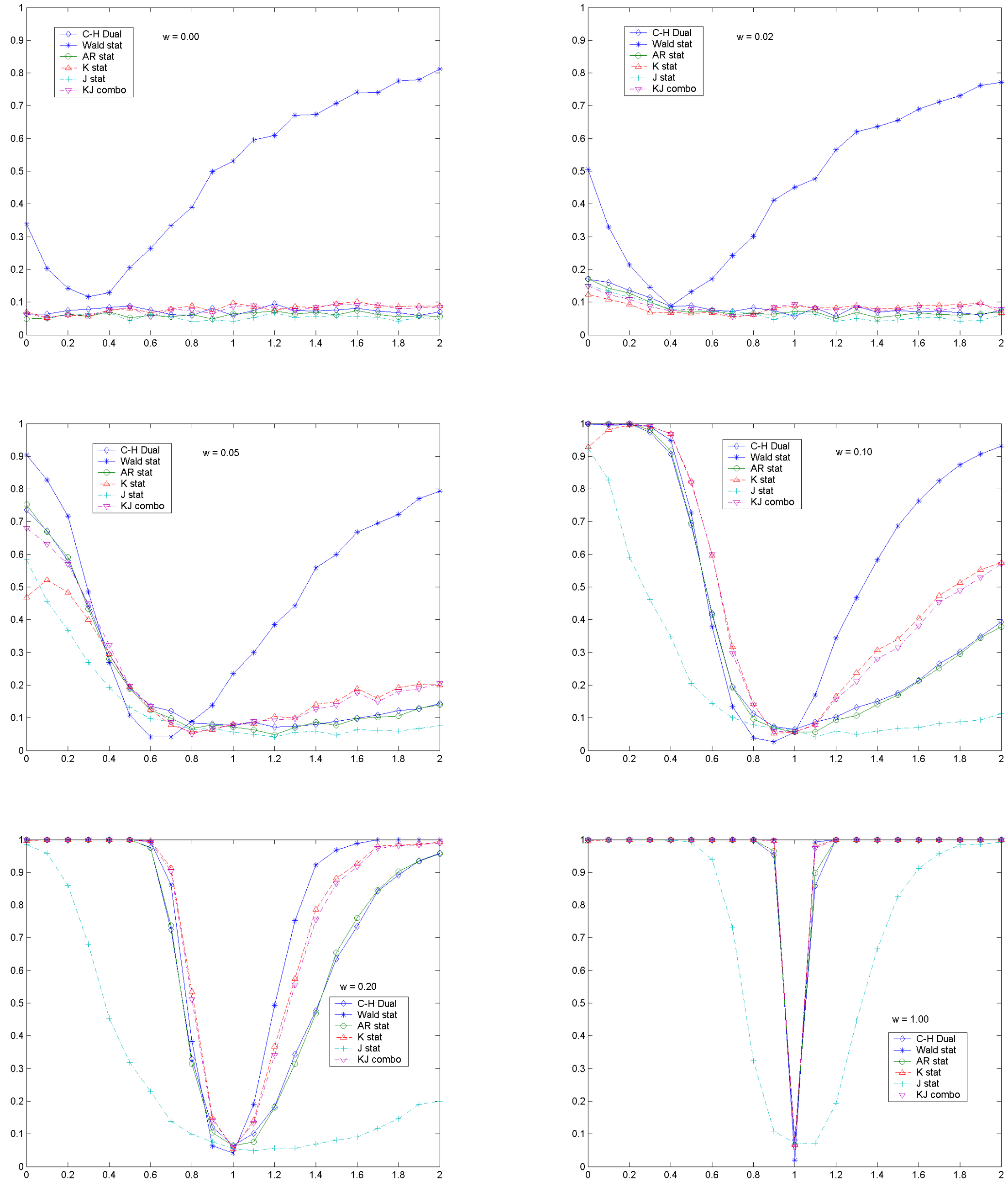
[18] The bandwidth choice issue is not addressed here because it is beyond the scope of this paper.

[19] $(\widehat{\alpha}(\tau), \widehat{\beta}(\tau)) = \arg\min_{\alpha,\beta} (\sum_{i=1}^{n} Q_{Xi}(K(\frac{D_i\alpha+\beta-\overline{Y}_i}{h_n}) - \tau))'(\frac{1}{n}\sum_{i=1}^{n} Q_{Xi}Q'_{Xi})^{-1}(\sum_{i=1}^{n} Q_{Xi}(K(\frac{D_i\alpha+\beta-\overline{Y}_i}{h}) - \tau))$, where $K(v)$ is a cdf type kernel and $Q_{Xi} = \begin{bmatrix} 1 & Z_i' \end{bmatrix}'$.

For $K(v)$, $\Phi(v)$, the cdf of the standard normal distribution was used (see McCurdy and Timmins (2001)).

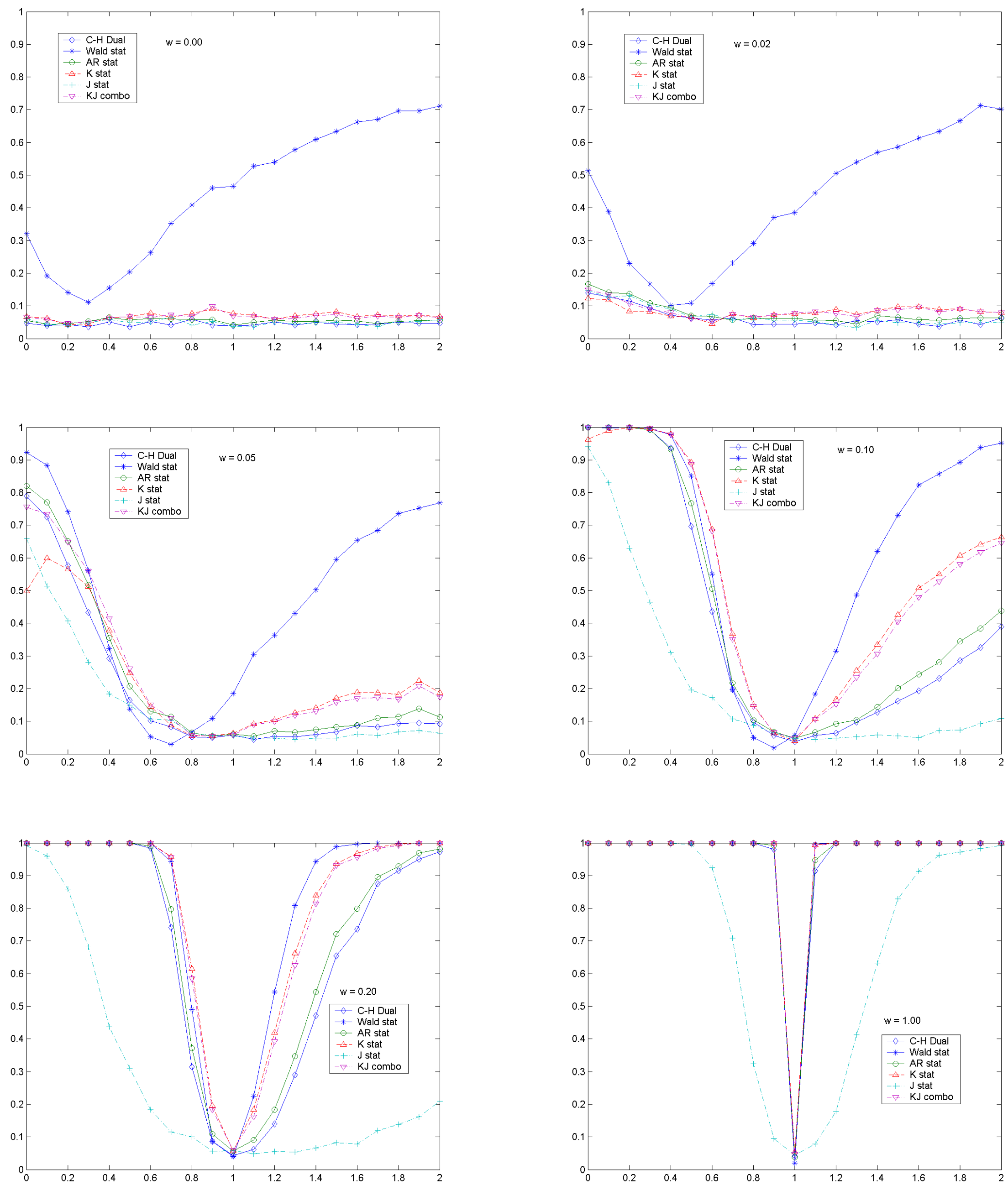
[20] Under $H_0 : \alpha(\tau) = \alpha_H$, run a quantile regression of $Y - D\alpha_H$ on $X$ and $Q_X$ and construct the Wald statistic testing the coefficients of $Q_X$ equal 0 (see Chernozhukov and Hansen (2004)).

Figure 1: Power Curves for Testing $\alpha(0.25) = 1$[21]



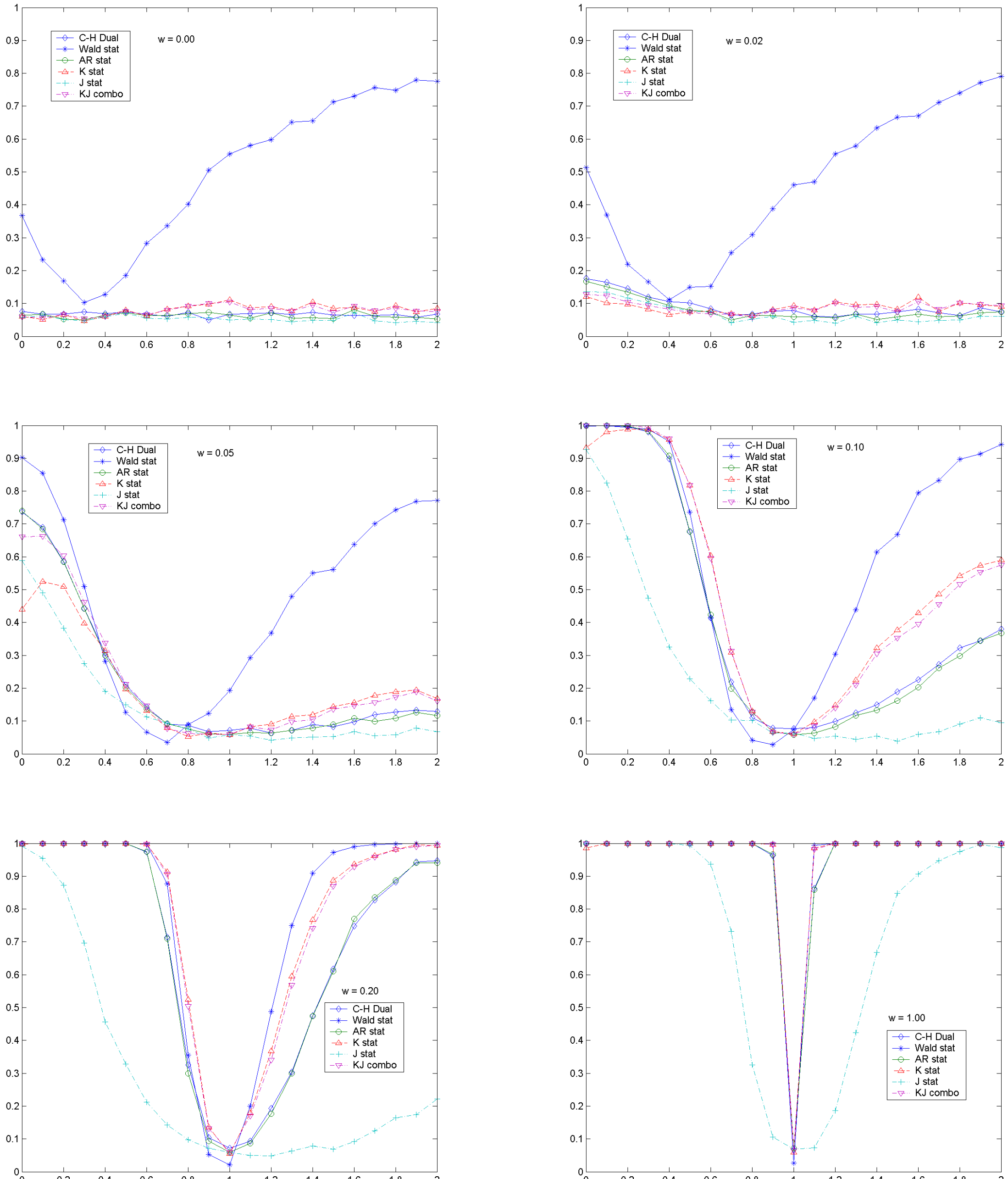

[21]The $x$ axis is $a_0$ and the $y$ axis is the level of rejection rates. The nominal size of these tests is 5%. For the $K$-$J$ combination, the experiments used 4% for part $K$ and 1% for part $J$.

Figure 2: Power Curves for Testing $\alpha(0.50) = 1$[22]




[22]The $x$ axis is $a_0$ and the $y$ axis is the level of rejection rates. The nominal size of these tests is 5%. For the $K$-$J$ combination, the experiments used 4% for part $K$ and 1% for part $J$.

Figure 3: Power Curves for Testing $\alpha(0.75) = 1$[23]




[23] The $x$ axis is $a_0$ and the $y$ axis is the level of rejection rates. The nominal size of these tests is 5%. For the $K$-$J$ combination, the experiments used 4% for part $K$ and 1% for part $J$.

# 7 Empirical Example: Returns to Schooling

This section considers an empirical example using the data analyzed by Angrist and Krueger (1991). Angrist and Krueger used birth quarter as an instrument for estimating the effects of compulsory schooling on earnings, which sparked much controversy about weak instruments. The sample used, consists of 329,509 men born between 1930-1939 from the 1980 census. This data set was used by Angrist, Imbens and Krueger (1999) and is accessible on the web via the Journal of Applied Econometrics. The dependent variable $Y$ is the log weekly wage, and the endogenous variable $D$ is the years of schooling. The exogenous variables ($X$) consist of 10 dummy variables indicating the year of birth. The instruments ($Z$) are the quarter of birth interacted with the year of birth. Hence, there are 30 instruments, 10 (included) exogenous variables, and a single endogenous variable. This is a simple specification of Angrist-Krueger, which does not include the state dummies and their interactions with the year of birth.

While Angrist et al. (1999) estimate the mean effect by the conventional 2SLS, LIML, and the jacknife IV method, Honore and Hu (2004), Lee (2004), and Chernozhukov and Hansen (2005b) consider the effects of schooling on various quantiles. Chernozhukov and Hansen (2005b) consider a different specification that includes the state dummies, and they use a different estimation method, which ignores the potential issue from weak instruments. Honore and Hu (2004) use the two-step $AR$ statistic with a sub-sample of white men. Lee (2004) investigates the same sample and the same specification as those considered here, but he uses a different method, which ignores the potential weak instrument issue. Although these previous studies use a different sample or specification, it is interesting to compare them with the results obtained here because the potential weak instrument issue is explicitly considered here.[24]

Lee (2004) and Chernozhukov and Hansen (2005b) found that education effects are larger at the lower end of the income distribution. It is, however, found that these results should be more carefully understood because the data are not informative for the lower quantile effects.

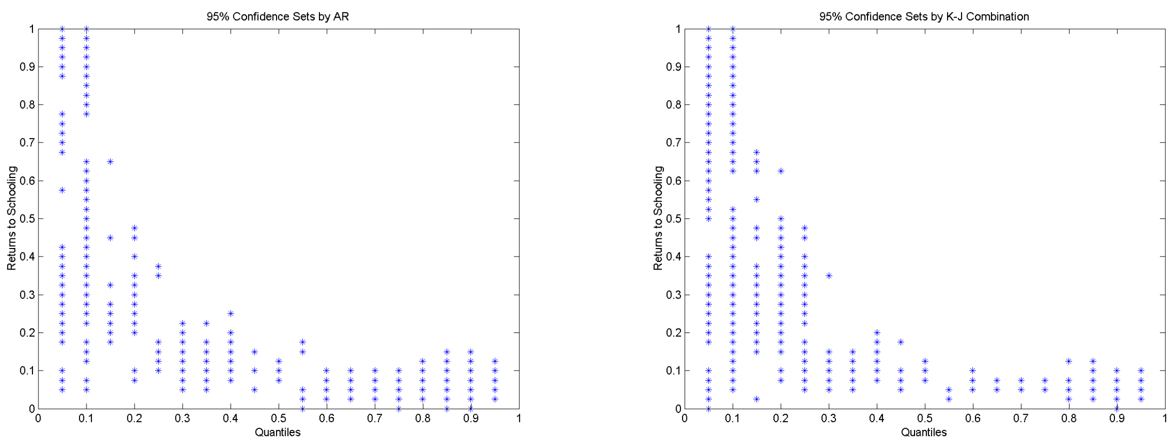
As a bandwidth choice, $h_n = n^{-\frac{1}{5}}$ was used.The pdf of the standard normal distribution was chosen as a kernel. The 95% confidence sets are provided in Figure 4. The returns to schooling for the $\tau^{th}$quantile are defined as the effect of the exogenous change of education on the $\tau^{th}$quantile

[24] Honore and Hu (2004) noted the weak instrument problem. However, they only considered the $25^{th}, 50^{th}$, and $75^{th}$ percentiles with the $AR$ statistic, which wastes many degrees of freedom.

response function of the potential log wage, which was denoted by $\alpha(\tau)$ in the previous sections. The values of $\alpha(\tau)$ that could not be rejected by the $AR$ test and the $K$-$J$ combination were collected to obtain the confidence sets. For the $K$-$J$ combination, 4% and 1% critical values were used for each $K$ and $J$ component, respectively.

The confidence sets found here are similar to those of Lee (2004) and Chernozhukov and Hansen (2005b) for the upper quantiles. However, the lower quantiles are quite different. The robust confidence sets for the lower quantiles are much wider than the results in those studies. This indicates that the weak instrument problem is concerned with the lower end of the income distribution and that the tight confidence sets for these quantiles are misleading. Although education effects on the upper quantiles seem to be quite constant, the effects on the lower quantiles still remain obscure. The confidence sets for the upper quantiles could be tightened by using the $K$-$J$ combination.

Figure 4: 95% Confidence Sets for Returns to Schooling[25]




[25] The exogenous variables are 10 dummies indicating the year of birth, and the instruments are 30 dummies that consist of the quarter of birth interacted with the year of birth. For the $K$-$J$ combination, 4% and 1% critical values were used for each $K$ and $J$ component, respectively.

# 8 Conclusion

We considered the linear instrumental quantile model proposed by Chernozhukov and Hansen (2001, 2005a, 2005b) and developed test procedures that are robust to the potential identification difficulty (Dufour (1997) and Stock and Wright (2000), among others). Specifically, we first considered a two-step $AR$ statistic and obtained a three-step orthogonal decomposition of the $AR$ statistic, where the null distribution of each component is robust to weak values of the Jacobian of the concentrated moment condition. Removing the identification assumption is particularly interesting in quantile models because (1) defending instruments is not as easy as in mean regression models, and (2) given the instruments, the identification depends on what quantile effects are examined. It is never clear for what quantile effects the given instruments are most informative. In particular, when the instruments are weak, the identification of some quantile effects can be relatively harder than that of others.

As an empirical example, returns to schooling using the Angrist-Krueger data were examined, and confidence sets for various quantile effects were obtained. Although returns to schooling for the upper quantiles seem to be quite constant, the robust confidence sets for the lower quantiles were far wider than those obtained by Lee (2004) and Chernozhukov and Hansen (2005b). This indicates that the given instruments are not informative for the lower quantiles, and thus education effects on the lower distribution of income still remain obscure. The confidence sets for the upper quantiles could also be tightened by using the decomposition of the $AR$ statistic.

As an alternative in the literature, a control function approach in a triangular system has also been suggested for endogenous quantile analysis (Imbens and Newey (2002), Chesher (2003), Ma and Koenker (2004), Lee (2004)). The advantage of the GMM approach is that the weak identification problem is well understood while there is little discussion on this topic in the control function approach. There is, however, an important limitation in the GMM approach: it only considers different quantiles of outcomes and does not explicitly model heterogeneous effects of instruments on an endogenous variable. In constrast, the triangular system directly models the first-stage equation and thus provides the better framework to discuss the idea of local instruments and identification pitfalls: the instruments do not affect a particular quantile of the endogenous variable although they do affect others. Therefore, identification, lack of identification, and inference in the triangular system are worthy of further study, and they are left for the future research.

# Appendix A: Lemma

First, we state the following lemma, which is used extensively.

**Lemma A** Suppose $g$ is a Borel measurable function on $\mathcal{R}^k$ such that

(i) $\int \mid g(u) \mid du < \infty$

(ii) $\parallel u \parallel^k \mid g(u) \mid \rightarrow 0$as $\parallel u \parallel \rightarrow 0$

(iii) $\sup \mid g(u) \mid < \infty$

Suppose that $f$ is another function on $\mathcal{R}^k$such that $\int \mid f(u) \mid du < \infty$.

Then, at every continuity point $x$ of $f$,we have $\int \frac{1}{h^k} g(\frac{z}{h}) f(x-z) dz \rightarrow f(x) \int g(u) du$ as $h \downarrow 0$.

If $f$ is uniformly continuous, then this convergence is also uniform.

For proof, see Pagan and Ullah (1999, p.362).

# Appendix B: Proof of Theorem 1

From lemma 1, we have $\begin{bmatrix} \sqrt{n}\widehat{m} \\ \sqrt{n}vec(\widehat{J}) - \sqrt{n}vec(J(\alpha(\tau), \beta(\tau))) \end{bmatrix}$

$\Rightarrow \begin{bmatrix} \widetilde{\Psi}_1 \\ vec(\widetilde{\Psi}_2) \end{bmatrix} \sim N(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \widetilde{V}_{11} & 0 \\ 0 & \widetilde{V}_{22} - \widetilde{V}_{21}\widetilde{V}_{11}^{-1}\widetilde{V}_{12} \end{bmatrix}$)as $n \rightarrow \infty$.

In particular, $\widetilde{\Psi}_1$and $\widetilde{\Psi}_2$are independent.

Hence, $S_n^{*\prime}(\widetilde{\Psi}_2' \widetilde{V}_{11}^{-1} \widetilde{\Psi}_2)^{-1} S_n^* \mid \widetilde{\Psi}_2 \xrightarrow{d} \chi^2(1)$, and since the asymptotic distribution does not depend on $\widetilde{\Psi}_2$, it is unconditional.

From $\sqrt{n}\widehat{J} \Rightarrow \widetilde{\Psi}_2$,we have $K_n^* \equiv S_n^{*\prime}((\sqrt{n}\widehat{J})' \widetilde{V}_{11}^{-1} (\sqrt{n}\widehat{J}))^{-1} S_n^* \xrightarrow{d} \chi^2(1+k)$.

(i) When $J(\alpha(\tau), \beta(\tau)) = 0$ : note that $\sqrt{n}\widehat{J} \Rightarrow \widetilde{\Psi}_2$.

Therefore, $S_n^* \equiv (\sqrt{n}\widehat{J})' \widetilde{V}_{11}^{-1} (\sqrt{n}\widehat{m}_J) \Rightarrow \widetilde{\Psi}_2' \widetilde{V}_{11}^{-1} \widetilde{\Psi}_1 \mid \widetilde{\Psi}_2 \sim N(0, \widetilde{\Psi}_2' \widetilde{V}_{11}^{-1} \widetilde{\Psi}_2)$.

Hence, $S_n^{*\prime}(\widetilde{\Psi}_2' \widetilde{V}_{11}^{-1} \widetilde{\Psi}_2)^{-1} S_n^* \mid \widetilde{\Psi}_2 \xrightarrow{d} \chi^2(1+k)$, and since the asymptotic distribution does not depend on $\widetilde{\Psi}_2$, it is unconditional.

From $\sqrt{n}\widehat{J} \Rightarrow \widetilde{\Psi}_2$, we have $K_n^* \equiv S_n^{*\prime}((\sqrt{n}\widehat{J})' \widetilde{V}_{11}^{-1} (\sqrt{n}\widehat{J}))^{-1} S_n^* \xrightarrow{d} \chi^2(1+k)$.

(ii) When $J(\alpha(\tau), \beta(\tau)) = C/\sqrt{n}$, where $C$ is of full rank: note that $\sqrt{n}\widehat{J} \Rightarrow \widetilde{\Psi}_2 + C$.

Therefore, $S_n^* \equiv (\sqrt{n}\widehat{J})' \widetilde{V}_{11}^{-1} (\sqrt{n}\widehat{m}_J) \Rightarrow (\widetilde{\Psi}_2 + C)' \widetilde{V}_{11}^{-1} \widetilde{\Psi}_1 \mid \widetilde{\Psi}_2 \sim N(0, (\widetilde{\Psi}_2 + C)' \widetilde{V}_{11}^{-1} (\widetilde{\Psi}_2 + C))$.

Hence, $S_n^{*\prime}((\widetilde{\Psi}_2 + C)' \widetilde{V}_{11}^{-1} (\widetilde{\Psi}_2 + C))^{-1} S_n^* \mid \widetilde{\Psi}_2 \xrightarrow{d} \chi^2(1+k)$, and since the asymptotic distribution does not depend on $\widetilde{\Psi}_2$, it is unconditional.

From $\sqrt{n}\widehat{J} \Rightarrow \widetilde{\Psi}_2 + C$, we have $K_n^* \equiv S_n^{*\prime}((\sqrt{n}\widehat{J})'\widetilde{V}_{11}^{-1}(\sqrt{n}\widehat{J}))^{-1}S_n^* \xrightarrow{d} \chi^2(1+k)$.

(iii) When $J(\alpha(\tau), \beta(\tau))$ is of full rank: note that $\widehat{J} \xrightarrow{p} J(\alpha(\tau), \beta(\tau))$.

Therefore, $\frac{1}{\sqrt{n}}S_n^* \equiv \widehat{J}'\widetilde{V}_{11}^{-1}(\sqrt{n}\widehat{m}_J) \Rightarrow J(\alpha(\tau), \beta(\tau))'\widetilde{V}_{11}^{-1}\widetilde{\Psi}_1 \sim N(0, J(\alpha(\tau), \beta(\tau))'\widetilde{V}_{11}^{-1}J(\alpha(\tau), \beta(\tau)))$.

Hence, $(\frac{1}{\sqrt{n}}S_n^*)'(J(\alpha(\tau), \beta(\tau))'\widetilde{V}_{11}^{-1}J(\alpha(\tau), \beta(\tau)))^{-1}(\frac{1}{\sqrt{n}}S_n^*) \xrightarrow{d} \chi^2(1+k)$.

From $\widehat{J} \xrightarrow{p} J(\alpha(\tau), \beta(\tau))$, we have $K_n^* \equiv S_n^{*\prime}((\sqrt{n}\widehat{J})'\widetilde{V}_{11}^{-1}(\sqrt{n}\widehat{J}))^{-1}S_n^* \xrightarrow{d} \chi^2(1+k)$. ■

# Appendix C: Proof of Lemma 2

Recall that $m_1(\alpha, \beta(\tau, \alpha)) = E(X(1\{\overline{Y} \leq D\alpha + X'\beta(\tau, \alpha)\} - \tau)) = 0$ by the definition of $\beta(\tau, \alpha)$ for each $\alpha$.

Letting $\epsilon \equiv \overline{Y} - D\alpha(\tau) - X'\beta(\tau, \alpha(\tau))$, we have $m_1(\alpha, \beta(\tau, \alpha)) = E(X(1\{\epsilon \leq D(\alpha - \alpha(\tau)) + X'(\beta(\tau, \alpha) - \beta(\tau, \alpha(\tau)))\} - \tau)) = 0$.

Note that we can rewrite this as $m_1(\alpha, \beta(\tau, \alpha)) = E(X(F_\epsilon(D(\alpha - \alpha(\tau)) + X'(\beta(\tau, \alpha) - \beta(\tau, \alpha(\tau)) \mid D, Z, X) - \tau)) = 0$due to the law of iterated expectation, where $F_\epsilon(\cdot \mid \cdot)$is the conditional distribution function of $\epsilon$.

Under assumption 4, we apply implicit function theorem to obtain

$$\frac{\partial \beta(\tau, \alpha)}{\partial \alpha} \mid_{\alpha=\alpha(\tau)} = -E(XX'f_\epsilon(0 \mid D, Z, X))^{-1}E(XDf_\epsilon(0 \mid D, Z, X))$$

Now, consider the concentrated moment function, $m_c(\alpha) \equiv m_2(\alpha, \beta(\tau, \alpha)) = E(W(1\{\overline{Y} \leq D\alpha + X'\beta(\tau, \alpha)\} - \tau)) = E(W(F_\epsilon(D(\alpha - \alpha(\tau)) + X'(\beta(\tau, \alpha) - \beta(\tau, \alpha(\tau)) \mid D, Z, X) - \tau))$. (We used the definition of $\epsilon$ and the law of iterated expectation.)

By the chain rule, we know that $\frac{\partial m_c}{\partial \alpha} \mid_{\alpha=\alpha(\tau)} = \frac{\partial m_2}{\partial \alpha} \mid_{\alpha=\alpha(\tau), \beta=\beta(\tau, \alpha(\tau))} + (\frac{\partial m_2}{\partial \beta'} \mid_{\alpha=\alpha(\tau), \beta=\beta(\tau, \alpha(\tau))})(\frac{\partial \beta(\tau, \alpha)}{\partial \alpha} \mid_{\alpha=\alpha(\tau)})$.

Therefore, we have

$$\begin{aligned}\frac{\partial m_c}{\partial \alpha} & \mid \quad {}_{\alpha=\alpha(\tau)} \\ &= E(WDf_\epsilon(0 \mid D, Z, X)) - E(WX'f_\epsilon(0 \mid D, Z, X))E(XX'f_\epsilon(0 \mid D, Z, X))^{-1}E(XDf_\epsilon(0 \mid D, Z, X)) \\ &= E(W - AX)Df_\epsilon(0 \mid D, Z, X))\end{aligned}$$

where $A \equiv E(WX'f_\epsilon(0 \mid D, Z, X))E(XX'f_\epsilon(0 \mid D, Z, X))^{-1}$.

## Appendix D: Proof of Lemma 3

First, note that the convergence of each element of $\widehat{\Gamma}_{n,h}$ is enough because $\widehat{\Gamma}_{n,h}$ is only finite dimensional.

Let $\widehat{\Gamma}^j_{n,h}$ be the $j^{th}$ element of $\widehat{\Gamma}_{n,h}$. Similarly, we will denote the $j^{th}$ element of $\Gamma$ as $\Gamma^j$.

Then, $E((\widehat{\Gamma}^j_{n,h} - \Gamma^j)^2) = Var(\widehat{\Gamma}^j_{n,h}) + (E(\widehat{\Gamma}^j_{n,h}) - \Gamma^j)^2$. (That is, $MSE = Variance + Bias^2$)

We will show that $Var(\widehat{\Gamma}^j_{n,h}) \to 0$ as $nh \to \infty$, $h \downarrow 0$ and that $(E(\widehat{\Gamma}^j_{n,h}) - \Gamma^j)^2 \to 0$ as $h \downarrow 0$.

First, $Var(\widehat{\Gamma}^j_{n,h}) = Var(\frac{1}{nh}\sum_{i=1}^n Q^j_i D_i k(\frac{-\epsilon_i}{h})) = \frac{1}{nh}\frac{1}{h}Var(Q^j_i D_i k(\frac{-\epsilon_i}{h}))$, where $Q^j_i$ is the $j^{th}$ element of $Q_i$.

Here, $\frac{1}{h}Var(Q^j_i D_i k(\frac{-\epsilon_i}{h})) = \frac{1}{h}\{E((Q^j_i D_i)^2 k(\frac{-\epsilon_i}{h})^2) - E(Q^j_i D_i k(\frac{-\epsilon_i}{h}))^2\}$.

Note that $\frac{1}{h}E((Q^j_i D_i)^2 k(\frac{-\epsilon_i}{h})^2) = E((Q^j_i D_i)^2 E(\frac{1}{h}k(\frac{-\epsilon_i}{h})^2 \mid D_i, Z_i, X_i)) = E((Q^j_i D_i)^2 \int \frac{1}{h}k(\frac{-s}{h})^2 f_\epsilon(s \mid D_i, Z_i, X_i)ds)$ by the law of iterated expectation.

Since $\int \frac{1}{h}k(\frac{-s}{h})^2 f_\epsilon(s \mid D_i, Z_i, X_i)ds \to f_\epsilon(0 \mid D_i, Z_i, X_i)\int k(v)^2 dv$ as $h \downarrow 0$ by lemma A, we know that $E((Q^j_i D_i)^2 \int \frac{1}{h}k(\frac{-s}{h})^2 f_\epsilon(s \mid D_i, Z_i, X_i)ds) \to E((Q^j_i D_i)^2 f_\epsilon(0 \mid D_i, Z_i, X_i))\int k(s)^2 ds$ as $h \downarrow 0$.

Similarly, $\frac{1}{h}E(Q^j_i D_i k(\frac{-\epsilon_i}{h}))^2 = h\{\frac{1}{h}E(Q^j_i D_i k(\frac{-\epsilon_i}{h}))\}^2 = h\{E(Q^j_i D_i \int \frac{1}{h}k(\frac{-s}{h}) f_\epsilon(s \mid D_i, Z_i, X_i)ds)\}^2 \to 0 \cdot \{E(Q^j_i D_i f_\epsilon(s \mid D_i, Z_i, X_i))\int k(s)ds\}^2 = 0$ as $h \downarrow 0$. (We used the law of iterated expectation and lemma A again.)

Hence we know that $\frac{1}{h}Var(Q^j_i D_i k(\frac{-\epsilon_i}{h})) = \frac{1}{h}\{E((Q^j_i D_i)^2 k(\frac{-\epsilon_i}{h})^2) - E(Q^j_i D_i k(\frac{-\epsilon_i}{h}))^2\} \to E((Q^j_i D_i)^2 f_\epsilon(0 \mid D_i, Z_i, X_i))\int k(s)^2 ds$ as $h \downarrow 0$.

Therefore, we conclude that $Var(\widehat{\Gamma}^j_{n,h}) = \frac{1}{nh}\frac{1}{h}Var(Q^j_i D_i k(\frac{-\epsilon_i}{h})) \to 0$ as $nh \to \infty$ and $h \downarrow 0$.

Next, consider the bias term.

Note that $E(\widehat{\Gamma}^j_{n,h}) = E(\frac{1}{h}Q^j_i D_i k(\frac{-\epsilon_i}{h})) = E(Q^j_i D_i \int \frac{1}{h}k(\frac{-s}{h}) f_\epsilon(s \mid D_i, Z_i, X_i)ds)$ by the law of iterated expectation.

Since $\int \frac{1}{h}k(\frac{-s}{h}) f_\epsilon(s \mid D_i, Z_i, X_i)ds \to f_\epsilon(0 \mid D_i, Z_i, X_i)\int k(v)dv$ as $h \downarrow 0$ by lemma A, we have $E(Q^j_i D_i \int \frac{1}{h}k(\frac{-s}{h}) f_\epsilon(s \mid D_i, Z_i, X_i)ds) \to E(Q^j_i D_i f_\epsilon(0 \mid D_i, Z_i, X_i))\int k(v)dv = \Gamma^j$ as $h \downarrow 0$.

Since the vaiance converges to 0 as $nh \to \infty$ , $h \downarrow 0$ and the bias goes to 0 as $h \downarrow 0$, we conclude that the mean squared error of $\widehat{\Gamma}^j_{n,h}$ converges to 0 as $nh \to \infty, h \downarrow 0$. Hence, $\widehat{\Gamma}^j_{n,h}$ converges in probability to $\Gamma^j$ as $nh \to \infty, h \downarrow 0$ by Chebyshev inequality. ■

## Appendix E: Proof of Lemma 4

Assuming the data are iid, $Var(\begin{bmatrix} \sqrt{n}\widehat{m}_c \\ \sqrt{nh}(\widehat{\Gamma}_{n,h} - E(\widehat{\Gamma}_{n,h})) \end{bmatrix}) = \begin{bmatrix} V & V_{12,h} \\ V_{21,h} & V_{22,h} \end{bmatrix}$ is given as follows.

(i) $V = E(W_iW_i'(1\{Y_i \leq D_i\alpha(\tau)\} - \tau)^2) - m_c(\alpha(\tau))m_c(\alpha(\tau))'$.

(ii) $V_{22,h} = \frac{1}{h}E(Q_iQ_i'D_i^2k^2(\frac{-\epsilon_i}{h})) - \frac{1}{h}E(Q_iD_ik(\frac{-\epsilon_i}{h}))E(Q_i'D_ik(\frac{-\epsilon_i}{h}))$.

Note that $\frac{1}{h}E(Q_iQ_i'D_i^2k^2(\frac{-\epsilon_i}{h})) = E(Q_iQ_i'D_i^2\int\frac{1}{h}k^2(\frac{-s}{h})f_\epsilon(s \mid D_i, Z_i, X_i)ds)$ by the law of iterated expectation.

Since $\int\frac{1}{h}k^2(\frac{-s}{h})f_\epsilon(s \mid D_i, Z_i, X_i)ds \to f_\epsilon(0 \mid D_i, Z_i, X_i)\int k(v)^2dv$ as $h \downarrow 0$ by lemma A, we have $E(Q_iQ_i'D_i^2\int\frac{1}{h}k^2(\frac{-s}{h})f_\epsilon(s \mid D_i, Z_i, X_i)ds) \to E(Q_iQ_i'D_i^2f_\epsilon(0 \mid D_i, Z_i, X_i))\int k(v)^2dv \equiv V_{22}$ as $h \downarrow 0$. For the second term, note that $\frac{1}{h}E(Q_iD_ik(\frac{-\epsilon_i}{h})) = E(Q_iD_i\int\frac{1}{h}k(\frac{-s}{h})f_\epsilon(s \mid D_i, Z_i, X_i)ds)$ by the law of iterated expectation and it converges to $E(Q_iD_if_\epsilon(0 \mid D_i, Z_i, X_i))\int k(v)dv = \Gamma$ as $h \downarrow 0$ by lemma A. Hence, $V_{22,h} = \frac{1}{h}E(Q_iQ_i'D_i^2k^2(\frac{-\epsilon_i}{h})) - h(\frac{1}{h}E(Q_iD_ik(\frac{-\epsilon_i}{h})))(\frac{1}{h}E(Q_iD_ik(\frac{-\epsilon_i}{h}))') \to V_{22} + 0\Gamma\Gamma' = V_{22}$ as $h \downarrow 0$.

(iii) $V_{12,h} = \frac{1}{\sqrt{h}}E(W_iQ_i'D_ik(\frac{-\epsilon_i}{h})(1\{\epsilon_i \leq 0\} - \tau)) - \frac{1}{\sqrt{h}}E(W_i(1\{\epsilon_i \leq 0\} - \tau))E(Q_iD_ik(\frac{-\epsilon_i}{h}))'$.

For the first term, $\frac{1}{\sqrt{h}}E(W_iQ_i'D_ik(\frac{-\epsilon_i}{h})(1\{\epsilon_i \leq 0\} - \tau)) = \sqrt{h}E(W_iQ_i'D_i\frac{1}{h}k(\frac{-\epsilon_i}{h})(1\{\epsilon_i \leq 0\} - \tau))$ $= \sqrt{h}E(W_iQ_i'D_i\int\frac{1}{h}k(\frac{-s}{h})(1\{s \leq 0\} - \tau)f_\epsilon(s \mid D_i, Z_i, X_i)ds) = \sqrt{h}E(W_iQ_i'D_i\int\frac{1}{h}k(\frac{-s}{h})(1\{\frac{-s}{h} \geq 0\} - \tau)f_\epsilon(s \mid D_i, Z_i, X_i)ds)$. The first equality is simply multiplying and dividing by $\sqrt{h}$. The second equality is by the law of iterated expectation and the third is because $1\{s \leq 0\} = 1\{\frac{-s}{h} \geq 0\}$ for all $s$. Then, considering $k(-z)(1\{-z \geq 0\} - \tau)$ as the function $g$ in lemma A, we know that $\sqrt{h}E(W_iQ_i'D_i\int\frac{1}{h}k(\frac{-s}{h})(1\{\frac{-s}{h} \geq 0\} - \tau)f_\epsilon(s \mid D_i, Z_i, X_i)ds) \to 0 \cdot E(W_iQ_i'D_if_\epsilon(0 \mid D_i, Z_i, X_i))\int k(v)(1\{v \geq 0\} - \tau)dv = 0$ as $h \downarrow 0$.

The second term is also negligible because $\frac{1}{\sqrt{h}}E(W_i(1\{\epsilon_i \leq 0\} - \tau))E(Q_iD_ik(\frac{-\epsilon_i}{h}))' = \sqrt{h}E(W_i(1\{\epsilon_i \leq 0\} - \tau))E(Q_iD_i\frac{1}{h}k(\frac{-\epsilon_i}{h}))' \to 0 \cdot E(W_i(1\{\epsilon_i \leq 0\} - \tau))E(Q_iD_if_\epsilon(0 \mid D_i, Z_i, X_i))'\int k(v)dv$ as $h \downarrow 0$. Here the first equality is multiplying and dividing by $\sqrt{h}$ and the convergence is again by the law of iterated expectation and lemma A.

Hence, we conclude that $V_{12,h} \to 0$ as $h \downarrow 0$.

Therefore, $\begin{bmatrix} V & V_{12,h} \\ V_{21,h} & V_{22,h} \end{bmatrix} \to \begin{bmatrix} V & 0 \\ 0 & V_{22} \end{bmatrix}$ as $h \downarrow 0$.

Recall that for any sequence $h_n > 0$ with $h_n \to 0, nh_n \to \infty$,

we have $\begin{bmatrix} V & V_{12,h_n} \\ V_{21,h_n} & V_{22,h_n} \end{bmatrix}^{-\frac{1}{2}} \begin{bmatrix} \sqrt{n}\widehat{m}_c \\ \sqrt{nh}(\widehat{\Gamma}_{n,h_n} - E(\widehat{\Gamma}_{n,h_n})) \end{bmatrix} \xrightarrow{d} N(0, I)$ as $n \to \infty$ by assumption 7.

Hence, we know that $\begin{bmatrix} V & 0 \\ 0 & V_{22} \end{bmatrix}^{-\frac{1}{2}} \begin{bmatrix} \sqrt{n}\widehat{m}_c \\ \sqrt{nh}(\widehat{\Gamma}_{n,h_n} - E(\widehat{\Gamma}_{n,h_n})) \end{bmatrix}$

$$= \left( \begin{bmatrix} V & 0 \\ 0 & V_{22} \end{bmatrix}^{-\frac{1}{2}} \begin{bmatrix} V & V_{12,h_n} \\ V_{21,h_n} & V_{22,h_n} \end{bmatrix}^{\frac{1}{2}} - I \right) \begin{bmatrix} V & V_{12,h_n} \\ V_{21,h_n} & V_{22,h_n} \end{bmatrix}^{-\frac{1}{2}} \begin{bmatrix} \sqrt{n}\widehat{m}_c \\ \sqrt{nh}(\widehat{\Gamma}_{n,h_n} - E(\widehat{\Gamma}_{n,h_n})) \end{bmatrix} +$$
$$\begin{bmatrix} V & V_{12,h_n} \\ V_{21,h_n} & V_{22,h_n} \end{bmatrix}^{-\frac{1}{2}} \begin{bmatrix} \sqrt{n}\widehat{m}_c \\ \sqrt{nh}(\widehat{\Gamma}_{n,h_n} - E(\widehat{\Gamma}_{n,h_n})) \end{bmatrix}$$
$$= o(1)O_p(1) + \begin{bmatrix} V & V_{12,h_n} \\ V_{21,h_n} & V_{22,h_n} \end{bmatrix}^{-\frac{1}{2}} \begin{bmatrix} \sqrt{n}\widehat{m}_c \\ \sqrt{nh}(\widehat{\Gamma} - E(\widehat{\Gamma})) \end{bmatrix} \xrightarrow{d} N(0, I) \text{ as } n \to \infty.$$

Therefore, we conclude that $\begin{bmatrix} \sqrt{n}\widehat{m}_c \\ \sqrt{nh}(\widehat{\Gamma}_{n,h_n} - E(\widehat{\Gamma}_{n,h_n})) \end{bmatrix} \xrightarrow{d} N(0, \begin{bmatrix} V & 0 \\ 0 & V_{22} \end{bmatrix})$ as $n \to \infty$.

Now, consider the limit of $\sqrt{nh_n}(E(\widehat{\Gamma}_{n,h_n}) - \Gamma)$.

Note that $\sqrt{nh_n}(E(\widehat{\Gamma}_{n,h_n}) - \Gamma) = \sqrt{nh_n}(E(\frac{1}{h_n}Q_iD_ik(\frac{-\epsilon_i}{h_n})) - E(Q_iD_if_\epsilon(0 \mid D_i, Z_i, X_i)))$ by definition of $\widehat{\Gamma}_{n,h_n}$, $\Gamma$.

It then follows that $\sqrt{nh_n}(E(\widehat{\Gamma}_{n,h_n}) - \Gamma) = \sqrt{nh_n}E(Q_iD_i(E(\frac{1}{h_n}k(\frac{-\epsilon_i}{h_n}) \mid D_i, Z_i, X_i) - f_\epsilon(0 \mid D_i, Z_i, X_i)))$ by the law of iterated expectation.

Here, $E(\frac{1}{h_n}k(\frac{-\epsilon_i}{h_n}) \mid D_i, Z_i, X_i) - f_\epsilon(0 \mid D_i, Z_i, X_i) = \int \frac{1}{h_n}k(\frac{-s}{h_n})f_\epsilon(s \mid D_i, Z_i, X_i)ds - f_\epsilon(0 \mid D_i, Z_i, X_i) \int k(s)ds = \int k(v)(f_\epsilon(-h_nv \mid D_i, Z_i, X_i) - f_\epsilon(0 \mid D_i, Z_i, X_i))dv$

$= \int k(v)(-f'_\epsilon(0 \mid D_i, Z_i, X_i)h_nv + \frac{1}{2}f''_\epsilon(v^* \mid D_i, Z_i, X_i)h_n^2v^2)dv$, where $v^*$ is between 0 and $-h_nv$.

Hence, $\sqrt{nh_n}(E(\widehat{\Gamma}_{n,h_n}) - \Gamma) = \sqrt{nh_n}h_nE(Q_iD_i(-f'_\epsilon(0 \mid D_i, Z_i, X_i)) \int vk(v)dv + \sqrt{nh_n}h_n^2\frac{1}{2}E(Q_iD_if''_\epsilon(v^* \mid D_i, Z_i, X_i)) \int v^2k(v)dv$

$$\to B = \begin{cases} c_1E(Q_iD_i(-f'_\epsilon(0 \mid D_i, Z_i, X_i)) \int vk(v)dv & \text{if } \sqrt{nh_n}h_n \to c_1 \\ c_2\frac{1}{2}E(Q_iD_if''_\epsilon(0 \mid D_i, Z_i, X_i)) \int v^2k(v)dv & \text{if } \sqrt{nh_n}h_n^2 \to c_2 \text{ and } \int vk(v)dv = 0 \end{cases}$$
as $n \to \infty$.

Finally, note that $\begin{bmatrix} \sqrt{n}\widehat{m}_c \\ \sqrt{nh_n}(\widehat{\Gamma}_{n,h_n} - \Gamma) \end{bmatrix} = \begin{bmatrix} \sqrt{n}\widehat{m}_c \\ \sqrt{nh_n}(\widehat{\Gamma}_{n,h_n} - E(\widehat{\Gamma}_{n,h_n})) \end{bmatrix} + \begin{bmatrix} 0 \\ \sqrt{nh_n}(E(\widehat{\Gamma}_{n,h_n}) - \Gamma) \end{bmatrix}$

■

# Appendix F: Proof of Theorem 2

Consider a sequence $h_n > 0$ with $h_n \to 0, nh_{n \to \infty}, \sqrt{nh_n}h_n \to c_1$ (or $\sqrt{nh_n}h_n^2 \to c_2$ and $\int vk(v)dv = 0$).

Then, from lemma 4, we have $\begin{bmatrix} \sqrt{n}\widehat{m}_c \\ \sqrt{nh_n}(\widehat{\Gamma}_{n,h_n} - \Gamma) \end{bmatrix} \Rightarrow \begin{bmatrix} \Psi_1 \\ \Psi_2 \end{bmatrix} \sim N(\begin{bmatrix} 0 \\ B \end{bmatrix}, \begin{bmatrix} V & 0 \\ 0 & V_{22} \end{bmatrix})$ as $n \to \infty$.

In particular, $\Psi_1$ and $\Psi_2$ are independent.

(i) When $\Gamma = 0$ or $\Gamma = C_1/\sqrt{n}:$ note that $\sqrt{nh_n}\widehat{\Gamma}_{n,h_n} \Rightarrow \Psi_2$. Therefore, $S_n \equiv (\sqrt{nh_n}\widehat{\Gamma}_{n,h_n})'V^{-1}(\sqrt{n}\widehat{m}_c) \Rightarrow \Psi_2'V^{-1}\Psi_1 \mid \Psi_2 \sim N(0, \Psi_2'V^{-1}\Psi_2)$.

Hence, $S_n'(\Psi_2'V^{-1}\Psi_2)^{-1}S_n \mid \Psi_2 \xrightarrow{d} \chi^2(1)$ and since the asymptotic distribution does not depend on $\Psi_2$, it is unconditional.

From $\sqrt{nh_n}\widehat{\Gamma}_{n,h_n} \Rightarrow \Psi_2$, we have $K_n \equiv S_n'((\sqrt{nh_n}\widehat{\Gamma}_{n,h_n})'V^{-1}(\sqrt{nh_n}\widehat{\Gamma}_{n,h_n}))^{-1}S_n \xrightarrow{d} \chi^2(1)$ as $n \to \infty$.

(ii) When $\Gamma = C_2/\sqrt{nh_n}$, where $C_2$ is of full rank: note that $\sqrt{nh_n}D_{n,h} \Rightarrow \Psi_2 + C_2$.

Therefore, $S_n \equiv (\sqrt{nh_n}\widehat{\Gamma}_{n,h_n})'V^{-1}(\sqrt{n}\widehat{m}_c) \Rightarrow (\Psi_2+C_2)'V^{-1}\Psi_1 \mid \Psi_2 \sim N(0, (\Psi_2+C_2)'V^{-1}(\Psi_2+C_2))$.

Hence, $S_n'((\Psi_2 + C_2)'V^{-1}(\Psi_2 + C_2))^{-1}S_n \mid \Psi_2 \xrightarrow{d} \chi^2(1)$ and since the asymptotic distribution does not depend on $\Psi_2$, it is unconditional.

From $\sqrt{nh_n}\widehat{\Gamma}_{n,h_n} \Rightarrow \Psi_2 + C_2$, we have $K_n \equiv S_n'((\sqrt{nh_n}\widehat{\Gamma}_{n,h_n})'V^{-1}(\sqrt{nh_n}\widehat{\Gamma}_{n,h_n}))^{-1}S_n \xrightarrow{d} \chi^2(1)$ as $n \to \infty$.

(iii) When $\Gamma$ is of full rank: note that $\widehat{\Gamma}_{n,h_n} \xrightarrow{p} \Gamma$. Therefore, $\frac{1}{\sqrt{nh_n}}S_n \equiv \widehat{\Gamma}'_{n,h_n}V^{-1}(\sqrt{n}\widehat{m}_c) \Rightarrow \Gamma'V^{-1}\Psi_1 \sim N(0, \Gamma'V^{-1}\Gamma)$.

Hence, $(\frac{1}{\sqrt{nh_n}}S_n)'(\Gamma'V^{-1}\Gamma)^{-1}(\frac{1}{\sqrt{nh_n}}S_n) \xrightarrow{d} \chi^2(1)$.

From $\widehat{\Gamma}_{n,h_n} \xrightarrow{p} \Gamma$, we have $K_n \equiv (\frac{1}{\sqrt{nh_n}}S_n)'(\widehat{\Gamma}'_{n,h_n}V^{-1}\widehat{\Gamma}_{n,h_n})^{-1}(\frac{1}{\sqrt{nh_n}}S_n) \xrightarrow{d} \chi^2(1)$ as $n \to \infty$. ■

## Appendix G: Proof of Lemma 5

For the sake of notational simplicity, we will write $\beta(\tau)$ and $\widehat{\beta}(\tau)$ for $\beta(\tau, \alpha(\tau))$ and $\widehat{\beta}(\tau, \alpha(\tau))$, respectively.

Note that

$$
\begin{aligned}
& G_{n,h_n}(\alpha(\tau), \beta) \\
= & \{G_{n,h_n}(\alpha(\tau), \beta) - G_{h_n}(\alpha(\tau), \beta)\} + G_{h_n}(\alpha(\tau), \beta) \\
= & G(\alpha(\tau), \beta) + \{G_{n,h_n}(\alpha(\tau), \beta) - G_{h_n}(\alpha(\tau), \beta)\} + \{G_{h_n}(\alpha(\tau), \beta) - G(\alpha(\tau), \beta)\} \\
= & H(\beta - \beta(\tau)) + o(\| \beta - \beta(\tau) \|) + \{G_{n,h_n}(\alpha(\tau), \beta) - G_{h_n}(\alpha(\tau), \beta)\} + \{G_{h_n}(\alpha(\tau), \beta) - G(\alpha(\tau), \beta)\} \\
= & H(\beta - \beta(\tau)) + G_{n,h_n}(\alpha(\tau), \beta(\tau)) - G_{n,h_n}(\alpha(\tau), \beta(\tau)) \\
+o( & \| \beta - \beta(\tau) \|) + \{G_{n,h_n}(\alpha(\tau), \beta) - G_{h_n}(\alpha(\tau), \beta)\} + \{G_{h_n}(\alpha(\tau), \beta) - G(\alpha(\tau), \beta)\} \\
= & L_{n,h_n}(\alpha(\tau), \beta) + \{G_{n,h_n}(\alpha(\tau), \beta) - G_{h_n}(\alpha(\tau), \beta)\} - \{G_{n,h_n}(\alpha(\tau), \beta(\tau)) - G_{h_n}(\alpha(\tau), \beta(\tau))\} \\
+o( & \| \beta - \beta(\tau) \|) + \{G_{h_n}(\alpha(\tau), \beta) - G(\alpha(\tau), \beta)\}
\end{aligned}
$$

where $L_{n,h_n}(\alpha(\tau), \beta) \equiv H(\beta - \beta(\tau)) + G_{n,h_n}(\alpha(\tau), \beta(\tau))$. (The first and second equalities are simply adding and subtracting. The third equality is by assumption 10. The fourth equality is adding and subtracting $G_{n,h_n}(\alpha(\tau), \beta(\tau))$. The last equality is by the definition of $L_{n,h_n}$ and the fact that $G_{h_n}(\alpha(\tau), \beta(\tau)) = 0$.) The intuition for this decomposition is that $G_{n,h_n}$ consists of the linear term, the stochastic equicontinuity term, the bias term, and the remainder.

Hence,

$$
\begin{aligned}
\| & \Upsilon_{n,h_n}(G_{n,h_n}(\alpha(\tau), \beta) - L_{n,h_n}(\alpha(\tau), \beta)) \| \qquad \text{(G.1)} \\
\leq & \| \Upsilon_{n,h_n}\{(G_{n,h_n}(\alpha(\tau), \beta) - G_{h_n}(\alpha(\tau), \beta)) - (G_{n,h_n}(\alpha(\tau), \beta(\tau)) - G_{h_n}(\alpha(\tau), \beta(\tau)))\} \| \\
+ \quad \| & \Upsilon_{n,h_n}\{G_{h_n}(\alpha(\tau), \beta) - G(\alpha(\tau), \beta)\} \| + \| \Upsilon_{n,h_n} o(\| \beta - \beta(\tau) \|) \| \\
= & \| \Upsilon_{n,h_n}\{(G_{n,h_n}(\alpha(\tau), \beta) - G_{h_n}(\alpha(\tau), \beta)) - (G_{n,h_n}(\alpha(\tau), \beta(\tau)) - G_{h_n}(\alpha(\tau), \beta(\tau)))\} \| \\
+ \quad \| & \Upsilon_{n,h_n}\{G_{h_n}(\alpha(\tau), \beta) - G(\alpha(\tau), \beta)\} \| + \left\| \begin{bmatrix} o(\sqrt{n} \| \beta - \beta(\tau) \|) \\ o(\sqrt{nh_n} \| \beta - \beta(\tau) \|) \end{bmatrix} \right\|
\end{aligned}
$$

Here, note that $\Upsilon_{n,h_n}\{G_{h_n}(\alpha(\tau), \beta) - G(\alpha(\tau), \beta)\} = \begin{bmatrix} 0 \\ B_{n,h_n}(\alpha(\tau), \beta) - B_{n,h_n}(\alpha(\tau), \beta(\tau)) \end{bmatrix}$. (See assumption 13.)

Therefore, for any $\delta_n \downarrow 0$, we have

$\sup_{\|\beta-\beta(\tau)\|\leq\delta_n} \| \Upsilon_{n,h_n}\{G_{h_n}(\alpha(\tau),\beta) - G(\alpha(\tau),\beta)\} \|$

$= \sup_{\|\beta-\beta(\tau)\|\leq\delta_n} \| B_{n,h_n}(\alpha(\tau),\beta) - B_{n,h_n}(\alpha(\tau),\beta(\tau)) \|$

$= \sup_{\|\beta-\beta(\tau)\|\leq\delta_n} \| B_{n,h_n}(\alpha(\tau),\beta) - B(\alpha(\tau),\beta) + B(\alpha(\tau),\beta) - B(\alpha(\tau),\beta(\tau)) + B(\alpha(\tau),\beta(\tau)) - B_{n,h_n}(\alpha(\tau),\beta(\tau)) \|$

$\leq \sup_{\|\beta-\beta(\tau)\|\leq\delta_n} \| B_{n,h_n}(\alpha(\tau),\beta) - B(\alpha(\tau),\beta) \| + \sup_{\|\beta-\beta(\tau)\|\leq\delta_n} \| B(\alpha(\tau),\beta) - B(\alpha(\tau),\beta(\tau)) \| + \sup_{\|\beta-\beta(\tau)\|\leq\delta_n} \| B(\alpha(\tau),\beta(\tau)) - B_{n,h_n}(\alpha(\tau),\beta(\tau)) \|$

$\to 0$ as $n \to \infty$. (By assumption 13.)

Hence, we know that $\| \Upsilon_{n,h_n}\{G_{h_n}(\alpha(\tau),\widehat{\beta}(\tau)) - G(\alpha(\tau),\widehat{\beta}(\tau))\} \| = o_p(1)$ with probability going to 1.

From this, the inequality (G.1), and assumptions 8 and 12, we know that

$$\| \Upsilon_{n,h_n}(G_{n,h}(\alpha(\tau),\widehat{\beta}(\tau)) - L_{n,h_n}(\alpha(\tau),\widehat{\beta}(\tau))) \| \leq o_p(1) + \left\| \begin{bmatrix} o(O_p(1)) \\ \sqrt{h_n}o(O_p(1)) \end{bmatrix} \right\| = o_p(1)$$

with probability going to 1.

Therefore, $\Upsilon_{n,h_n}(G_{n,h_n}(\alpha(\tau),\widehat{\beta}(\tau)) = \Upsilon_{n,h_n}L_{n,h_n}(\alpha(\tau),\widehat{\beta}(\tau))) + o_p(1) = \Upsilon_{n,h_n}H(\widehat{\beta}(\tau) - \beta(\tau)) + \Upsilon_{n,h_n}G_{n,h_n}(\alpha(\tau),\beta(\tau)) + o_p(1)$ with probability going to 1.

It then follows that $\Upsilon_{n,h_n}(G_{n,h_n}(\alpha(\tau),\widehat{\beta}(\tau)) = \Upsilon_{n,h_n}G_{n,h_n}(\alpha(\tau),\beta(\tau)) + \begin{bmatrix} H_1\sqrt{n}(\widehat{\beta}(\tau) - \beta(\tau)) \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \sqrt{h_n}H_2\sqrt{n}(\widehat{\beta}(\tau) - \beta(\tau)) \end{bmatrix} + o_p(1)$.

Hence, we conclude that $\Upsilon_{n,h_n}(G_{n,h_n}(\alpha(\tau),\widehat{\beta}(\tau)) = \Upsilon_{n,h_n}G_{n,h_n}(\alpha(\tau),\beta(\tau)) + \begin{bmatrix} H_1\sqrt{n}(\widehat{\beta}(\tau) - \beta(\tau)) \\ 0 \end{bmatrix} + o_p(1)$ because $\sqrt{h_n}H_2\sqrt{n}(\widehat{\beta}(\tau) - \beta(\tau)) = o(1)O_p(1) = o_p(1)$.

■

## Appendix H: Proof of Lemma 6

For the sake of notational simplicity, we will write $\beta(\tau)$ and $\widehat{\beta}(\tau)$ for $\beta(\tau,\alpha(\tau))$ and $\widehat{\beta}(\tau,\alpha(\tau))$, respectively.

Note that $\Upsilon_{n,h}G_{n,h}(\alpha(\tau),\beta(\tau))+\begin{bmatrix} H_1\frac{1}{\sqrt{n}}\sum_{i=1}^n \Xi_i \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^n A_i \\ \frac{1}{\sqrt{nh}}\sum_{i=1}^n B_i \end{bmatrix}$ where $A_i \equiv \{W_i(1\{\epsilon_i \leq 0\}-\tau)+H_1\Xi_i\}$, $B_i \equiv (Q_iD_ik(\frac{-\epsilon_i}{h})-E(Q_iD_ik(\frac{-\epsilon_i}{h})))$.

Assuming the *iid* data, we can find the variance $V^*, V^*_{21,h}=V^{*\prime}_{12,h}$ and $V^*_{22,h}$. We then follow the proof af lemma 4 to show $V^*_{21,h}=V^{*\prime}_{12,h}\to 0$ and $V^*_{22,h}\to V^*_{22}$ as $h\downarrow 0$ by using lemma A.

Hence, for a sequence $h_n>0$ with $h_n\to 0, nh_n\to\infty$, as $n\to\infty$, we have $\begin{bmatrix} V^* & V^*_{12,h_n} \\ V^*_{21,h_n} & V^*_{22,h_n} \end{bmatrix} \to \begin{bmatrix} V^* & 0 \\ 0 & V^*_{22} \end{bmatrix}$ as $n\to\infty$.

Recall that we have $\begin{bmatrix} V^* & V^*_{12,h_n} \\ V^*_{21,h_n} & V^*_{22,h_n} \end{bmatrix}^{-\frac{1}{2}} \{\Upsilon_{n,h_n}G_{n,h_n}(\alpha(\tau),\beta(\tau))+\begin{bmatrix} H_1\frac{1}{\sqrt{n}}\sum_{i=1}^n \Xi_i \\ 0 \end{bmatrix})\} \xrightarrow{d} N(0,I)$ as $n\to\infty$ from assumption 11.

Therefore, we have

$$
\begin{aligned}
&\begin{bmatrix} V^* & 0 \\ 0 & V^*_{22} \end{bmatrix}^{-\frac{1}{2}} \Upsilon_{n,h_n}(G_{n,h_n}(\alpha(\tau),\widehat{\beta}(\tau)) \\
=&\begin{bmatrix} V^* & 0 \\ 0 & V^*_{22} \end{bmatrix}^{-\frac{1}{2}} \{\Upsilon_{n,h_n}G_{n,h_n}(\alpha(\tau),\beta(\tau))+\begin{bmatrix} H_1\frac{1}{\sqrt{n}}\sum_{i=1}^n \Xi_i \\ 0 \end{bmatrix}\}+o_p(1) \\
=&\begin{bmatrix} V^* & V^*_{12,h_n} \\ V^*_{21,h_n} & V^*_{22,h_n} \end{bmatrix}^{-\frac{1}{2}} \{\Upsilon_{n,h_n}G_{n,h_n}(\alpha(\tau),\beta(\tau))+\begin{bmatrix} H_1\frac{1}{\sqrt{n}}\sum_{i=1}^n \Xi_i \\ 0 \end{bmatrix})\}+o_p(1) \\
&\xrightarrow{d} N(0,I) \text{ as } n \text{ goes to infinity.}
\end{aligned}
$$

(The first equality is from lemma 5 and assumption 8. The second equality is because $\begin{bmatrix} V^* & V^*_{12,h_n} \\ V^*_{21,h_n} & V^*_{22,h_n} \end{bmatrix} \to \begin{bmatrix} V^* & 0 \\ 0 & V^*_{22} \end{bmatrix}$ as $n\to\infty$. Also, refer to the proof of lemma 4. The last convergence is from assumption 11.)

Hence, we conclude that $\Upsilon_{n,h_n}(G_{n,h_n}(\alpha(\tau),\widehat{\beta}(\tau)) \xrightarrow{d} N(0, \begin{bmatrix} V^* & 0 \\ 0 & V_{22}^* \end{bmatrix})$ as $n \to \infty$.

Now, we consider the convergence of the bias term, $\sqrt{nh_n}(E(\widehat{\Gamma}_{n,h_n}(\alpha(\tau),\beta(\tau))) - \Gamma)$, where $\Gamma$ denotes $\Gamma(\alpha(\tau),\beta(\tau))$.

Recall that $\sqrt{nh_n}(E(\widehat{\Gamma}_{n,h_n}(\alpha(\tau),\beta(\tau))) - \Gamma) \to B(\alpha(\tau),\beta(\tau))$ from assumption 13.

Finally, note that $\begin{bmatrix} \sqrt{n}\widehat{m}(\alpha(\tau),\widehat{\beta}(\tau)) \\ \sqrt{nh_n}(\widehat{\Gamma}_{n,h_n}(\alpha(\tau),\widehat{\beta}(\tau)) - \Gamma) \end{bmatrix}$

$= \Upsilon_{n,h_n}(G_{n,h_n}(\alpha(\tau),\widehat{\beta}(\tau)) + \begin{bmatrix} 0 \\ \sqrt{nh_n}(E(\widehat{\Gamma}_{n,h_n}(\alpha(\tau),\beta(\tau))) - \Gamma) \end{bmatrix}$. ■

## Appendix I: Using $\widehat{A}_{n,h_n}$ in place of the true $A$

All we need to show is that we can obtain lemma 6 by using $\widehat{A}_{n,h_n}$ instead of the true $A$. (That is, using $\widehat{Q}_i \equiv W_i - \widehat{A}_{n,h_n} X_i$ instead of $Q_i \equiv W_i - AX_i$.) The purpose of this appendix is to provide the assumptions for the counterpart of lemma 6 by using $\widehat{A}_{n,h_n}$ in place of the true $A$. We will state the assumptions we need in a manner parallel to section 5. As in section 5, $h$ denotes a positive number and $h_n$ denotes a sequential bandwidth choice.

Recall that $A = E(WX'f_\epsilon(0 \mid D, Z, X))E(XX'f_\epsilon(0 \mid D, Z, X))^{-1}$ and
$\widehat{A}_{n,h_n} = (\frac{1}{nh_n}\sum_{i=1}^n W_i X_i' k(\frac{D_i\alpha(\tau)+X_i'\widehat{\beta}(\tau,\alpha(\tau))-\overline{Y}_i}{h_n}))(\frac{1}{nh_n}\sum_{i=1}^n X_i X_i' k(\frac{D_i\alpha(\tau)+X_i'\widehat{\beta}(\tau,\alpha(\tau))-\overline{Y}_i}{h_n}))^{-1}$.

**Lemma I-1** $vec(\widehat{A}_{n,h_n}) - vec(A)$

$$= \begin{bmatrix} (\frac{1}{nh_n}\sum_{i=1}^n X_i X_i'\widehat{k}_i)^{-1} \otimes I) & -((\frac{1}{nh_n}\sum_{i=1}^n X_i X_i'\widehat{k}_i)^{-1} \otimes A) \end{bmatrix} \begin{bmatrix} \frac{1}{nh_n}\sum_{i=1}^n vec(W_i X_i')\widehat{k}_i - E(vec(W_i X_i')f_\epsilon) \\ \frac{1}{nh_n}\sum_{i=1}^n vec(X_i X_i')\widehat{k}_i - E(vec(X_i X_i')f_\epsilon) \end{bmatrix},$$

where $f_e \equiv f_\epsilon(0 \mid D_i, Z_i, X_i)$ and $\widehat{k}_i \equiv k(\frac{D_i\alpha(\tau)+X_i'\widehat{\beta}(\tau,\alpha(\tau))-\overline{Y}_i}{h_n})$.

**Proof.** At the end of this appendix.

From this lemma, we know that the asymptotic distribution of $\sqrt{nh_n}(vec(\widehat{A}_{n,h_n}) - vec(A))$ can be deduced from $\sqrt{nh_n}\begin{bmatrix} \frac{1}{nh_n}\sum_{i=1}^n vec(W_i X_i')\widehat{k}_i - E(vec(W_i X_i')f_\epsilon) \\ \frac{1}{nh_n}\sum_{i=1}^n vec(X_i X_i')\widehat{k}_i - E(vec(X_i X_i')f_\epsilon) \end{bmatrix}$.

First, consider $\begin{bmatrix} \widehat{m}_c(\alpha(\tau),\beta) \\ \widehat{\Gamma}^*_{n,h}(\alpha(\tau),\beta) \end{bmatrix}$, where

$$\widehat{\Gamma}^*_{n,h}(\alpha(\tau),\beta) \equiv \begin{bmatrix} \widehat{\Gamma}_{n,h}(\alpha(\tau),\beta) \\ \widehat{\Theta}_{n,h}(\alpha(\tau),\beta) \\ \widehat{\Delta}_{n,h}(\alpha(\tau),\beta) \end{bmatrix} \equiv \begin{bmatrix} \frac{1}{nh}\sum_{i=1}^n Q_i D_i k(\frac{D_i\alpha(\tau)+X_i'\beta-\overline{Y}_i}{h}) \\ \frac{1}{nh_n}\sum_{i=1}^n vec(W_i X_i') k(\frac{D_i\alpha(\tau)+X_i'\beta-\overline{Y}_i}{h}) \\ \frac{1}{nh_n}\sum_{i=1}^n vec(X_i X_i') k(\frac{D_i\alpha(\tau)+X_i'\beta-\overline{Y}_i}{h}) \end{bmatrix} \quad \text{(I.1)}$$

Let $\Upsilon^*_{n,h} \equiv \begin{bmatrix} \sqrt{n}I & 0 \\ 0 & \sqrt{nh}I \end{bmatrix}$ which is conformable with $\begin{bmatrix} \widehat{m}_c(\alpha(\tau),\beta) \\ \widehat{\Gamma}^*_{n,h}(\alpha(\tau),\beta) \end{bmatrix}$.

**Definition I-1** *Define* $G^*_{n,h}(\alpha(\tau),\beta) \equiv \begin{bmatrix} \widehat{m}_c(\alpha(\tau),\beta) \\ \widehat{\Gamma}^*_{n,h}(\alpha(\tau),\beta) - E(\widehat{\Gamma}^*_{n,h}(\alpha(\tau),\beta(\tau,\alpha(\tau)))) \end{bmatrix}$
*and* $G^*_h(\alpha(\tau),\beta) \equiv E(G^*_{n,h}(\alpha(\tau),\beta)) \equiv \begin{bmatrix} m_c(\alpha(\tau),\beta) \\ E(\widehat{\Gamma}^*_{n,h}(\alpha(\tau),\beta)) - E(\widehat{\Gamma}^*_{n,h}(\alpha(\tau),\beta(\tau,\alpha(\tau)))) \end{bmatrix}$.

The following assumptions correspond to assumption 9, 10, 11, 12, and 13 in section 5.

**Assumption I-1** $G_h^*(\alpha(\tau),\beta) \to G^*(\alpha(\tau),\beta) \equiv \begin{bmatrix} m_c(\alpha(\tau),\beta) \\ \Gamma^*(\alpha(\tau),\beta) - \Gamma^*(\alpha(\tau),\beta(\tau,\alpha(\tau))) \end{bmatrix}$ uniformly, at least locally around $\beta(\tau,\alpha(\tau))$ as $h \downarrow 0$. Note that $\Gamma^*(\alpha(\tau),\beta) \equiv \begin{bmatrix} \Gamma(\alpha(\tau),\beta) \\ \Theta(\alpha(\tau),\beta) \\ \Delta(\alpha(\tau),\beta) \end{bmatrix}$, where $\Gamma(\alpha(\tau),\beta)$ is the same as assumption 9 and $\Theta(\alpha(\tau),\beta(\tau,\alpha(\tau))) = E(vec(W_iX_i')f_\epsilon)$, $\Delta(\alpha(\tau),\beta(\tau,\alpha(\tau))) = E(vec(X_iX_i')f_\epsilon)$.

**Assumption I-2** $G^*(\alpha(\tau),\beta)$ can be linearized with respect to $\beta$ around $\beta(\tau,\alpha(\tau))$.
That is, $G^*(\alpha(\tau),\beta) = H^*(\beta - \beta(\tau,\alpha(\tau))) + o(\| \beta - \beta(\tau,\alpha(\tau)) \|)$ for some $H^*$.
We will write $H^* = \begin{bmatrix} H_1 \\ H_2^* \end{bmatrix}$ where $H_1$ and $H_2^*$ are from the top and bottom parts of $G^*$. Note that $H_1$ is the same as assumption 10.

**Assumption I-3** Suppose
$Var(\Upsilon_{n,h}^* G_{n,h}^*(\alpha(\tau),\beta(\tau,\alpha(\tau))) + \begin{bmatrix} H_1 \frac{1}{\sqrt{n}} \sum_{i=1}^n \Xi_i \\ 0 \end{bmatrix}) = \begin{bmatrix} V^* & V_{12,h}^{**} \\ V_{21,h}^{**} & V_{22,h}^{**} \end{bmatrix}$, where $V^*$ is the same as assumption 11.
Then, for any sequence $h_n > 0$ with $h_n \to 0, nh_n \to \infty$, we have

$$\begin{bmatrix} V^* & V_{12,h_n}^{**} \\ V_{21,h_n}^{**} & V_{22,h_n}^{**} \end{bmatrix}^{-\frac{1}{2}} \{\Upsilon_{n,h_n}^* G_{n,h_n}^*(\alpha(\tau),\beta(\tau,\alpha(\tau))) + \begin{bmatrix} H_1 \frac{1}{\sqrt{n}} \sum_{i=1}^n \Xi_i \\ 0 \end{bmatrix})\} \xrightarrow{d} N(0,I) \text{ as } n \to \infty.$$

**Assumption I-4** For any $\delta_n \downarrow 0$ and for any sequence $h_n > 0$ with $h_n \to 0, nh_n \to \infty$, we have
$\sup_{\|\beta-\beta(\tau,\alpha(\tau))\|<\delta_n} \| \Upsilon_{n,h_n}^*(G_{n,h_n}^*(\alpha(\tau),\beta) - G_{h_n}^*(\alpha(\tau),\beta))$
$-\Upsilon_{n,h_n}^*(G_{n,h_n}^*(\alpha(\tau),\beta(\tau,\alpha(\tau))) - G_{h_n}^*(\alpha(\tau),\beta(\tau,\alpha(\tau))) \| = o_p(1)$ as $n \to \infty$.

**Assumption I-5** For any seqeunce $h_n > 0$ with $h_n \to 0, nh_n \to \infty$ and $\sqrt{nh_n}h_n \to c_1$ (or $\sqrt{nh_n}h_n^2 \to c_2$ if $\int vk(v)dv = 0$), $B_{n,h_n}^*(\alpha(\tau),\beta) \equiv \sqrt{nh_n}(E(\widehat{\Gamma}_{n,h_n}^*(\alpha(\tau),\beta)) - \Gamma^*(\alpha(\tau),\beta)) \to B^*(\alpha(\tau),\beta)$ uniformly in $\beta$, at least locally around $\beta(\tau,\alpha(\tau))$ as $n \to \infty$.
We also assume that $B^*(\alpha(\tau),\beta)$ and $\Gamma^*(\alpha(\tau),\beta)$ are continuous at $\beta = \beta(\tau,\alpha(\tau))$. Note that $B^*(\alpha(\tau),\beta) \equiv \begin{bmatrix} B(\alpha(\tau),\beta) \\ B_\Theta(\alpha(\tau),\beta) \\ B_\Delta(\alpha(\tau),\beta) \end{bmatrix}$, where $B(\alpha(\tau),\beta)$ is the same as assumption 13.

These correspond to the assumptions used in section 5. We are only considering the longer bottom part with the same top part of $G_{n,h}, G_h, G$.

Now, we obtain the following lemmas.

**Lemma I-2** Under these assumptions, for any sequence $h_n > 0$ with $h_n \to 0, nh_n \to \infty$, we have

$$\Upsilon^*_{n,h_n}(G^*_{n,h_n}(\alpha(\tau), \widehat{\beta}(\tau, \alpha(\tau))) = \Upsilon^*_{n,h_n} G^*_{n,h_n}(\alpha(\tau), \beta(\tau, \alpha(\tau))) + \begin{bmatrix} H_1\sqrt{n}(\widehat{\beta}(\tau, \alpha(\tau)) - \beta(\tau, \alpha(\tau))) \\ 0 \end{bmatrix}$$

$+o_p(1)$.

**Proof.** Follow the proof of lemma 5.

**Lemma I-3** As $h \downarrow 0$, we have $V^{**}_{21,h} = V^{**\prime}_{12,h} \to 0$ and $V^{**}_{22,h} \to V^{**}_{22}$.

In particular, for any sequence $h_n > 0$ with $h_n \to 0, nh_n \to \infty$,and $\sqrt{nh_n}h_n \to c_1$(or $\sqrt{nh_n}h_n^2 \to c_2$ and $\int vk(v)dv = 0$),we have

$$\begin{bmatrix} \sqrt{n}\widehat{m}_c(\alpha(\tau), \widehat{\beta}(\tau, \alpha(\tau))) \\ \sqrt{nh_n}(\widehat{\Gamma}^*_{n,h_n}(\alpha(\tau), \widehat{\beta}(\tau, \alpha(\tau))) - \Gamma^*(\alpha(\tau), \beta(\tau, \alpha(\tau)))) \end{bmatrix} \xrightarrow{d} N(\begin{bmatrix} 0 \\ B^*(\alpha(\tau), \beta(\tau, \alpha(\tau))) \end{bmatrix}, \begin{bmatrix} V^* & 0 \\ 0 & V^{**}_{22} \end{bmatrix})$$

as $n \to \infty$.

**Proof.** Follow the proof of lemma 6.

From lemma I-3, we have
$$\begin{bmatrix} \sqrt{n}\widehat{m}_c(\alpha(\tau), \widehat{\beta}(\tau, \alpha(\tau))) \\ \sqrt{nh_n}(\widehat{\Gamma}_{n,h_n}(\alpha(\tau), \widehat{\beta}(\tau, \alpha(\tau))) - \Gamma(\alpha(\tau), \beta(\tau, \alpha(\tau)))) \\ \sqrt{nh_n}(\widehat{\Theta}_{n,h}(\alpha(\tau), \widehat{\beta}(\tau, \alpha(\tau))) - E(vec(W_i X_i')f_\epsilon)) \\ \sqrt{nh_n}(\widehat{\Delta}_{n,h}(\alpha(\tau), \widehat{\beta}(\tau, \alpha(\tau))) - E(vec(X_i X_i')f_\epsilon)) \end{bmatrix}$$

$$\xrightarrow{d} N(\begin{bmatrix} 0 \\ B(\alpha(\tau), \beta(\tau, \alpha(\tau))) \\ B_\Theta(\alpha(\tau), \beta(\tau, \alpha(\tau))) \\ B_\Delta(\alpha(\tau), \beta(\tau, \alpha(\tau))) \end{bmatrix}, \begin{bmatrix} V^* & 0 \quad 0 \quad 0 \\ 0 & \\ 0 & V^{**}_{22} \\ 0 & \end{bmatrix}) \text{ as } n \to \infty.$$

Therefore, by lemma I-1, we have

$$\begin{bmatrix} \sqrt{n}\widehat{m}_c(\alpha(\tau), \widehat{\beta}(\tau, \alpha(\tau))) \\ \sqrt{nh_n}(\widehat{\Gamma}_{n,h_n}(\alpha(\tau), \widehat{\beta}(\tau, \alpha(\tau))) - \Gamma(\alpha(\tau), \beta(\tau, \alpha(\tau)))) \\ \sqrt{nh_n}(vec(\widehat{A}_{n,h_n}) - vec(A)) \end{bmatrix} \tag{I.2}$$

$$\xrightarrow{d} N(\begin{bmatrix} 0 \\ B(\alpha(\tau), \beta(\tau, \alpha(\tau))) \\ \kappa_1 B_\Theta(\alpha(\tau), \beta(\tau, \alpha(\tau))) + \kappa_2 B_\Delta(\alpha(\tau), \beta(\tau, \alpha(\tau))) \end{bmatrix}, \begin{bmatrix} V^* & 0 & 0 \\ 0 & V^*_{22} & V^*_{23} \\ 0 & V^*_{32} & V^*_{33} \end{bmatrix})$$

as $n \to \infty$, where $\begin{bmatrix} V^*_{22} & V^*_{23} \\ V^*_{32} & V^*_{33} \end{bmatrix}$ is defined from $V^{**}_{22}, \kappa_1$ and $\kappa_2$ and

$$\begin{bmatrix} \kappa_1 & \kappa_2 \end{bmatrix} \equiv p\lim_{n\to\infty} \begin{bmatrix} (\frac{1}{nh_n}\sum_{i=1}^n X_i X_i' \widehat{k}_i)^{-1} \otimes I) & \vdots & -((\frac{1}{nh_n}\sum_{i=1}^n X_i X_i' \widehat{k}_i)^{-1} \otimes A) \end{bmatrix}.$$

Hence, we obtain the following lemma, which is the counterpart of lemma 6, by using $\widehat{A}_{n,h_n}$ instead of the true $A$.

**Lemma I-4** Consider a sequence $h_n$ as lemma J-3.

Let $\widehat{\widehat{\Gamma}}_{n,h_n}(\alpha(\tau),\widehat{\beta}(\tau,\alpha(\tau))) \equiv \frac{1}{nh_n}\sum_{i=1}^n \widehat{Q}_i D_i k(\frac{D_i\alpha(\tau)+X_i'\widehat{\beta}(\tau,\alpha(\tau))-\overline{Y}_i}{h})$ with $\widehat{Q}_i = W_i - \widehat{A}_{n,h_n} X_i$ .

Then, we have $\begin{bmatrix} \sqrt{n}\widehat{m}_c(\alpha(\tau),\widehat{\beta}(\tau,\alpha(\tau))) \\ \sqrt{nh_n}(\widehat{\widehat{\Gamma}}_{n,h_n}(\alpha(\tau),\widehat{\beta}(\tau,\alpha(\tau))) - \Gamma(\alpha(\tau),\beta(\tau,\alpha(\tau)))) \end{bmatrix} \xrightarrow{d} N(\begin{bmatrix} 0 \\ \mathbb{B} \end{bmatrix}, \begin{bmatrix} V^* & 0 \\ 0 & \mathbb{V}_{22} \end{bmatrix})$ as $n \to \infty$ for some $\mathbb{B}$ and $\mathbb{V}_{22}$.

**Proof.** Note that $\widehat{\widehat{\Gamma}}_{n,h_n}(\alpha(\tau),\widehat{\beta}(\tau,\alpha(\tau)))$ $= \widehat{\Gamma}_{n,h_n}(\alpha(\tau),\widehat{\beta}(\tau,\alpha(\tau))) + (A-\widehat{A}_{n,h_n})\frac{1}{nh_n}\sum_{i=1}^n X_i D_i k(\frac{D_i\alpha(\tau)+X_i'\widehat{\beta}(\tau,\alpha(\tau))-\overline{Y}_i}{h})$. Then, it results from (I.2).

From lemma I-4, we can immediately obtain the counterpart of theorem 3 with $\widehat{A}_{n,h_n}$ in place of the true $A$.

Lastly, we provide the proof of lemma I-1.

**Proof of Lemma I-1**

We will write $\widehat{k}_i$ and $f_\epsilon$ for $k(\frac{D_i\alpha(\tau)+X_i'\widehat{\beta}(\tau,\alpha(\tau))-\overline{Y}_i}{h_n})$ and $f_\epsilon(0 \mid D,Z,X)$ for the sake of simplicity.

First, note that $\widehat{A}_{n,h_n} - A$

$= (\frac{1}{nh_n}\sum_{i=1}^n W_i X_i'\widehat{k}_i)(\frac{1}{nh_n}\sum_{i=1}^n X_i X_i'\widehat{k}_i)^{-1} - E(WX'f_\epsilon)E(XX'f_\epsilon)^{-1}$

$= (\frac{1}{nh_n}\sum_{i=1}^n W_i X_i'\widehat{k}_i)(\frac{1}{nh_n}\sum_{i=1}^n X_i X_i'\widehat{k}_i)^{-1}$
$-E(WX'f_\epsilon)\{E(XX'f_\epsilon)^{-1} - (\frac{1}{nh_n}\sum_{i=1}^n X_i X_i'\widehat{k}_i)^{-1} + (\frac{1}{nh_n}\sum_{i=1}^n X_i X_i'\widehat{k}_i)^{-1}\}$

$= \{\frac{1}{nh_n}\sum_{i=1}^n W_i X_i'\widehat{k}_i - E(WX'f_\epsilon)\}(\frac{1}{nh_n}\sum_{i=1}^n X_i X_i'\widehat{k}_i)^{-1}$
$-E(WX'f_\epsilon)\{E(XX'f_\epsilon)^{-1} - (\frac{1}{nh_n}\sum_{i=1}^n X_i X_i'\widehat{k}_i)^{-1}\}$

$= \{\frac{1}{nh_n}\sum_{i=1}^n W_i X_i'\widehat{k}_i - E(WX'f_\epsilon)\}(\frac{1}{nh_n}\sum_{i=1}^n X_i X_i'\widehat{k}_i)^{-1}$
$-E(WX'f_\epsilon)E(XX'f_\epsilon)^{-1}\{\frac{1}{nh_n}\sum_{i=1}^n X_i X_i'\widehat{k}_i - E(XX'f_\epsilon)\}(\frac{1}{nh_n}\sum_{i=1}^n X_i X_i'\widehat{k}_i)^{-1}$

$= \{\frac{1}{nh_n}\sum_{i=1}^n W_i X_i'\widehat{k}_i - E(WX'f_\epsilon)\}(\frac{1}{nh_n}\sum_{i=1}^n X_i X_i'\widehat{k}_i)^{-1}$
$-A\{\frac{1}{nh_n}\sum_{i=1}^n X_i X_i'\widehat{k}_i - E(XX'f_\epsilon)\}(\frac{1}{nh_n}\sum_{i=1}^n X_i X_i'\widehat{k}_i)^{-1}$ since $E(WX'f_\epsilon)E(XX'f_\epsilon)^{-1} = A$.

Then, take the *vec*-operator.

## Appendix J: Estimation of $V^*$

For the sake of notational simplicity, we will write $\beta(\tau)$ and $\widehat{\beta}(\tau)$ for $\beta(\tau,\alpha(\tau))$ and $\widehat{\beta}(\tau,\alpha(\tau))$, respectively.

Recall that $V^* = Var(\frac{1}{\sqrt{n}}\sum_{i=1}^n W_i(1\{\overline{Y}_i \leq D_i\alpha(\tau) + X_i'\beta(\tau)\} - \tau) + H_1\frac{1}{\sqrt{n}}\sum_{i=1}^n \Xi_i)$, where $H_1 = \frac{\partial m_c(\alpha(\tau),\beta)}{\partial \beta'}\mid_{\beta=\beta(\tau)} = E(W_i X_i' f_\epsilon(0 \mid D_i, Z_i, X_i))$.

Recall that $\sqrt{n}(\widehat{\beta}(\tau) - \beta(\tau)) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \Xi_i + o_p(1)$, where $\widehat{\beta}(\tau)$ is a GMM estimator from $m_1(\alpha(\tau),\beta) = E(X(1\{\overline{Y}_i - D_i\alpha(\tau) \leq X_i'\beta\} - \tau)) = 0$ when $\beta = \beta(\tau)$. Therefore, $\Xi_i$ is actually given by $-F^{-1}X_i(1\{\overline{Y}_i - D_i\alpha(\tau) \leq X_i'\beta(\tau)\} - \tau)$, where $F \equiv \frac{\partial m_1}{\partial \beta'}\mid_{\beta=\beta(\tau)} = E(X_i X_i' f_\epsilon(0 \mid D_i, Z_i, X_i))$ (see e.g., Pakes and Pollard (1989), Newey and McFadden (1994)).

Therefore, letting $I_i^\tau \equiv 1\{\overline{Y}_i \leq D_i\alpha(\tau) + X_i'\beta(\tau)\} - \tau$, we have

$$\begin{aligned} V^* &= Var(\frac{1}{\sqrt{n}}\sum_{i=1}^n (W_i I_i^\tau + H_1\Xi_i)) \\ &= Var(W_i I_i^\tau + H_1\Xi_i) \\ &= Var(W_i I_i^\tau - H_1 F^{-1} X_i I_i^\tau) \end{aligned}$$

Here, $H_1$ is consistently estimated by $\widehat{H}_1 \equiv \frac{1}{nh_n}\sum_{i=1}^n W_i X_i k(\frac{D_i\alpha(\tau) + X_i'\beta(\tau) - \overline{Y}_i}{h_n})$.
$F$ can also be estimated by $\widehat{F} \equiv \frac{1}{nh_n}\sum_{i=1}^n X_i X_i' k(\frac{D_i\alpha(\tau) + X_i'\beta(\tau) - \overline{Y}_i}{h_n})$.
Lastly, the unknown $\beta(\tau)$ can be replaced by its consistent estimator $\widehat{\beta}(\tau)$.

The given variance can then be estimated by the straightforward sample variance.

# Appendix K: Derivation of (6.2)

First, note that $Z_i \perp \begin{bmatrix} \beta(U_i) \\ V_i \end{bmatrix}$ and $\begin{bmatrix} \beta(U_i) \\ V_i \end{bmatrix} \sim N(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix})$, from which we have $\epsilon_i = \beta(U_i) - \beta(\tau, \alpha(\tau)) \mid V_i, Z_i \sim N(\rho V - \beta(\tau, \alpha(\tau)), 1 - \rho^2)$.

We then have $f_\epsilon(0 \mid Z_i, V_i) = (2\pi)^{-\frac{1}{2}}(1-\rho^2)^{-\frac{1}{2}} \exp(-\frac{1}{2(1-\rho^2)}(\rho V_i - \beta(\tau))^2)$.

Hence, we have the following.

$$
\begin{aligned}
E(Z_i D_i f_\epsilon(0 \quad &| \quad Z_i, V_i)) \\
&= E(Z_i(\gamma_1 + Z_i'\gamma_2 + V_i) f_\epsilon(0 \mid Z_i, V_i)) \\
&= (2\pi)^{-\frac{1}{2}}(1-\rho^2)^{-\frac{1}{2}} E(Z_i(\gamma_1 + Z_i'\gamma_2 + V_i)\exp(-\frac{1}{2(1-\rho^2)}(\rho V_i - \beta(\tau, \alpha(\tau)))^2)) \\
&= (2\pi)^{-\frac{1}{2}}(1-\rho^2)^{-\frac{1}{2}}(E(Z_i)\gamma_1 + E(Z_i Z_i')\gamma_2) E(\exp(-\frac{1}{2(1-\rho^2)}(\rho V_i - \beta(\tau, \alpha(\tau)))^2)) \\
&\quad + (2\pi)^{-\frac{1}{2}}(1-\rho^2)^{-\frac{1}{2}} E(Z_i) E(V_i \exp(-\frac{1}{2(1-\rho^2)}(\rho V_i - \beta(\tau, \alpha(\tau)))))
\end{aligned}
$$

Here, $E(\exp(-\frac{1}{2(1-\rho^2)}(\rho V_i - \beta(\tau, \alpha(\tau)))^2)) = \frac{1}{\sqrt{2\pi}} \int \exp(-\frac{1}{2(1-\rho^2)}(\rho v - \beta(\tau, \alpha(\tau)))^2) \exp(-\frac{v^2}{2}) dv = \sqrt{1-\rho^2} \exp(-\frac{\beta(\tau, \alpha(\tau))^2}{2})$.

Also, $E(V_i \exp(-\frac{1}{2(1-\rho^2)}(\rho V_i - \beta(\tau, \alpha(\tau)))^2)) = \frac{1}{\sqrt{2\pi}} \int v \exp(-\frac{1}{2(1-\rho^2)}(\rho v - \beta(\tau, \alpha(\tau)))^2) \exp(-\frac{v^2}{2}) dv = \beta(\tau, \alpha(\tau))\rho\sqrt{1-\rho^2} \exp(-\frac{\beta(\tau, \alpha(\tau))^2}{2})$.

Therefore,

$$
\begin{aligned}
E(Z_i D_i f_\epsilon(0 \quad &| \quad Z_i, V_i)) \qquad\qquad (K.1) \\
&= (2\pi)^{-\frac{1}{2}} \exp(-\frac{\beta(\tau, \alpha(\tau))^2}{2})(E(Z_i)\gamma_1 + E(Z_i Z_i')\gamma_2 + E(Z_i)\beta(\tau, \alpha(\tau))\rho)
\end{aligned}
$$

Similarly, we can obtain

$$
\begin{aligned}
E(Z_i f_\epsilon(0 \quad &| \quad Z_i, V_i)) \qquad\qquad (K.2) \\
&= (2\pi)^{-\frac{1}{2}}(1-\rho^2)^{-\frac{1}{2}} E(Z_i \exp(-\frac{1}{2(1-\rho^2)}(\rho V_i - \beta(\tau, \alpha(\tau)))^2)) \\
&= (2\pi)^{-\frac{1}{2}} \exp(-\frac{\beta(\tau, \alpha(\tau))^2}{2}) E(Z_i)
\end{aligned}
$$

$$
\begin{aligned}
&E(Df_\epsilon(0 \quad | \quad Z_i, V_i)) && \text{(K.3)}\\
&= (2\pi)^{-\frac{1}{2}}(1-\rho^2)^{-\frac{1}{2}}E((\gamma_1 + Z_i'\gamma_2 + V_i)\exp(-\frac{1}{2(1-\rho^2)}(\rho V_i - \beta(\tau,\alpha(\tau)))^2))\\
&= (2\pi)^{-\frac{1}{2}}\exp(-\frac{\beta(\tau,\alpha(\tau))^2}{2})(\gamma_1 + E(Z_i)'\gamma_2 + \beta(\tau,\alpha(\tau))\rho)
\end{aligned}
$$

$$
\begin{aligned}
&E(f_\epsilon(0 \quad | \quad Z_i, V_i)) && \text{(K.4)}\\
&= (2\pi)^{-\frac{1}{2}}\exp(-\frac{\beta(\tau,\alpha(\tau))^2}{2})
\end{aligned}
$$

Noting that $X = 1, Q_Z = Z$, we know from K.1, K.2, K.3, K.4 that

$$
\begin{aligned}
\Gamma &= E(ZDf_\epsilon(0 \mid V, Z)) - E(Zf_\epsilon(0 \mid V, Z))E(f_\epsilon(0 \mid V, Z))^{-1}E(Df_\epsilon(0 \mid V, Z))\\
&= (2\pi)^{-\frac{1}{2}}\exp(-\frac{\beta(\tau,\alpha(\tau))^2}{2})(E(Z_iZ_i') - E(Z_i)E(Z_i)')\gamma_2\\
&= (2\pi)^{-\frac{1}{2}}\exp(-\frac{\beta(\tau,\alpha(\tau))^2}{2})Var(Z_i)\gamma_2
\end{aligned}
$$

# References


[1] Abadie, A. J. Angrist and G. Imbens (2002): "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," Econometrica 70, 91-117

[2] Abadie, A. (1995): "Changes in Spanish Labor Income Structure During 1980's: A Quantile Regression Approach," CEMFI Working Paper, Madrid, Spain

[3] Amemiya, T. (1982): "Two Stage Absolute Deviations Estimators," Econometrica 50, 689-712

[4] Anderson, T.W. and H. Rubin (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," Annals of Mathematical Statistics 20, 46-63

[5] Andrews, D. (1994): "Empirical Process Methods in Econometrics," Handbook of Econometrics V4, Ch.37, 2248-2294

[6] Angrist, J.D. and A.B. Krueger (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?," The Quarterly Journal of Economics 106, 979-1014

[7] Angrist, J.D., G.W. Imbens and A.B. Krueger (1999): "Jacknife Instrumental Variables Estimation," Journal of Applied Econometrics 14, 57-67

[8] Buchinsky, M. (1994): "Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression," Econometrica 62, 405-458

[9] Buchinsky, M. (1998): "Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research," Journal of Human Resources 33, 88-126

[10] Chen, L-A and S. Portnoy (1996): "Two-Stage Regression Quantiles and Two-Stage Trimmed Least Squares Estimators for Structural Equation Models," Comm. Stat.-Theory and Method 25, 1005-1032

[11] Chernozhukov, V. and C. Hansen (2001): "An IV Model of Quantile Treatment Effects," MIT Working Paper

[12] Chernozhukov, V. and C. Hansen (2004): "Instrumental Quantile Regression," MIT Working Paper, Revision of Chernozhukov et al (2001)

[13] Chernozhukov, V. and C. Hansen (2005a): "An IV Model of Quantile Treatment Effects," Econometrica 73, 245-261

[14] Chernozhukov, V. and C. Hansen (2005b): "Instrumental Quantile Regression Inference for Structural and Treatment Effect Models," Journal of Econometrics, Forthcoming

[15] Chesher, A. (2003): "Identification in Nonseparable Models," Econometrica 71, 1405-1441

[16] Dufour, J-M (1997): "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," Econometrica 65, 1365-1388

[17] Hahn, J. and J. Hausman (2002): "A New Specification Test for the Validity of Instrumental Variables," Econometrica 70, 163-189

[18] Honore, B. and L. Hu (2004): "On the Performance of Some Robust Instrumental Variables Estimators," Journal of Business and Economic Statistics 22, 30-39

[19] Horowitz, J.L. (1992): "A Smoothed Maximum Score Estimator for the Binary Response Model," Econometrica 60, 505-531

[20] Imbens, G. and W. Newey (2002): "Identification and Estimation of Triangular Simultaneous Equation Models without Additivity," NBER Working Paper

[21] Kleibergen, F. (2002): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," Econometrica 70, 1781-1803

[22] Kleibergen, F. (2004): "Pivotal Statistics for Testing Subsets of Structural Parameters in the IV Regression Model," Review of Economics and Statistics 86, 418-423

[23] Kleibergen, F. (2005): "Testing Parameters in GMM without Assuming that They are Identified," Econometrica 73, 1103-1123

[24] Koenker R. and G. Basset Jr. (1978): "Regression Quantile," Econometrica 46, 33-50

[25] Koenker R. and K. Hallock (2000): "Quantile Regression: An Introduction," University of Illinois at Urbana-Champaign, mimeo

[26] Lee, S. (2004): "Endogeneity in Quantile Regression Models: A Control Function Approach," University College London Working Paper

[27] Ma, L. and R. Koenker (2004): "Quantile Regression Methods for Recursive Structural Equation Models," University of Illinois at Urbana-Champaign Working Paper

[28] MaCurdy T. and C. Timmins (2001): "Bounding the Influence of Attrition on Intertemporal Wage Variation in the NLSY," Department of Economics, Duke University, Working Paper

[29] Moreira, M. (2003): "A Conditional Likelihood Ratio Test for Structural Models," Econometrica 71, 1027-2048

[30] Newey, W. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," Handbook of Econometrics V4, Ch.36, 2113-2148

[31] Pagan, A. and A. Ullah (1999): "Nonparametric Econometrics," Cambridge University Press

[32] Pakes, A and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," Econometrica 57, 1027-1057

[33] Staiger, J. and J. Stock (1997): "Instrumental Variables Regression with Weak Instruments," Econometrica 65, 557-586

[34] Stock, J. and J. Wright (2000): "GMM with Weak Identification," Econometrica 68, 1055-1096

[35] Stock, J. J. Wright and M. Yogo (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," Journal of Business and Economic Statistics 20, 518-529

[36] van der Vaart, A.W. and J.A. Wellner (1996): "Weak Convergence and Empirical Processes," Springer-Verlag New York, Inc.

[37] Wright, J. (2002): "Testing the Null of Identification in GMM," Federal Reserve System, International Finance Discussion Papers Number 732